Quality Assurance & Testing for the Crop BMP Dataset: FDACS_UFGA8201_peanut.94.xlsx

John/Jane Doe, Univ. Florida

2026-01-25

1. Introduction

In preparing datasets based on the BMP data template, researchers need to check that their data are entered correctly, that all expected data are included, that the dataset is formatted as intended, and that all variables are correctly defined, including units of measurement. This report contains output from the R script "FL_Crop_BMP_QA_QC_single_dataset.Rmd", which conducts a series of quality control (QC) tests on a dataset prior to submitting the data to a repository or funding agency. Users are strongly urged to run this tool throughout dataset preparation rather than a single run upon a presumed "final" dataset. It is far easier to correct errors as an experiment is executed than waiting until all data acquisition and analyses are completed. Because this report includes guidance on possible sources of errors and emphasizes quality testing as a continuous process, use of the tool combines elements of quality assurance and quality control, hence "QA/QC". The R script reviews crop BMP datasets according to the "four C's", whereby a dataset is:

1. *Correct*: The values are accurate within the expected range of measurement error. We emphasize that the main error-checking should be done as a part of the normal data management pipeline prior to loading into the BMP template.
2. *Complete*: The dataset is complete enough to enable further analysis without researchers having to seek guidance on how the crop was grown, weather conditions, etc.
3. *Coherent*: Identifiers (keys) used to link data across sheets are used consistently.
4. *Compatible*: By linking the BMP terminology to the ICASA standards, we expect that datasets can be used with a wide range of tools including artificial intelligence, machine learning and either simulation or statistical models. Information on the ICASA standards is available at the ICASA GitHub site.

This document (the exported PDF) is produced by running the knit command within RStudio. It alternates between text (such as this section) and blocks of output from R. Users who are familiar with R and R Markdown should feel free to modify the markdown file as needed.

Issues we have encountered in test datasets include:

- Incorrect formatting of dates.
- Use of variables that include non-standard characters (e.g. "?").
- Introduction of new variables in unexpected positions.

We try to resolve these issues as they appear. Unresolved issues and suggestions for improvements should be reported at the "Issues" section of the GitHub site: https://github.com/DSSAT/Florida_Crop_BMP_Datasets

1.1. Checking that the file is read as expected

We first list all sheets in the file FDACS_UFGA8201_peanut.94.xlsx. The list includes sheets that are defined but have no data.

```
 [1] "START HERE"                "Terminology"
 [3] "List of sheets and keys"   "M1. Experiments"
 [5] "M2. Sites"                 "M3. Experimental Design"
 [7] "E1. Treatments"            "E2. Fields"
 [9] "E3. Plots"                 "E4. Crop Information"
[11] "E5. Planting"              "E6. Irrigation"
[13] "E7. Fertilizer"            "E9. Tillage"
[15] "E8. Organic Amendments"    "E10. Chemical Applications"
[17] "E11. Harvest"              "E12. Preplant Soil"
[19] "O1. Analysis Methods"      "O2. Yield Summary"
[21] "O3. Crop Growth"           "O4. Crop Health"
[23] "O5. Soil Surface Properties" "O6. Soil Layer Properties"
[25] "O7. Water"                 "S1. Soil Metadata"
[27] "S2. Soil Layer Properties" "W1. Weather Station Metadata"
[29] "W2. Daily Weather Data"    "Z1. Dictionary Metadata"
[31] "Z2. Dictionary Observations" "Z3. Dictionary Soils Weather"
```

2.0. Correct and Complete?: Summarizing the Content of Individual Sheets

Summaries are generated for the contents of each sheet except for the first three sheets, which contain instructions, and the last three, which contain the dictionaries. If sufficient numeric data are present, box plots are created for any numeric variables, including management levels.

Results for each sheet should be checked to make sure they match expectations for all variables. The QA/QC tool is *not* shoudl never serve as the primary means of detecting incorrect values. We assume the researchers have already conducted extensive data validation.

2.1. Summaries for each sheet (Tabular summaries first, then box plots of numeric variables).

If numeric data appear in tables of frequencies, this means the data for the variable has been interpreted as text (character sting). This can arise if there are any non-numeric values such as "." in the original data. Be sure to inspect cells in the column below any problem variables in case a character has inadvertently been entered below the intended data.

Depending on the amount of data in the sheets, the corresponding group of box plots may appear after the summary of the next sheet (i.e., the box plots will be slightly out of order).

Start processing of:M1___Experiments

Table 1: M1___Experiments

	Variable	Value	Frequency
Experiment name	Experiment name	Peanut cultivars x four N rates	1
Experiment ID	Experiment ID	UFGA8201	1
Research data owner	Research data owner	Selamat and Gardner	1
Institutional data owner	Institutional data owner	University of Florida	1
Publication journal & volume	Publication journal & volum	Agronomy J 77:862-867	1
Link to document	Link to document	https: //acsess.onlinelibrary.wiley.com/doi/ab	1

END Processing M1___Experiments

Start processing of:M2___Sites

Table 2: M2___Sites

	Variable	Value	Frequency
Site	Site	UFGA	1
Local name for experiment site	Local name for experiment s	Agronomy Farm	1
State	State	FL	1
County	County	Alachua	1
Town or other	Town or other	Gainesville	1

END Processing M2___Sites

Start processing of:M3___Experimental_Design

Table 3: M3___Experimental_Design

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	1
Site	Site	UFGA	1
Treatment structure	Treatment structure	RCBD	1
Type of experiment	Type of experiment	Station experiment	1
Main effect 1	Main effect 1	Cultivar	1
Main effect 2	Main effect 2	Nitrogen	1

END Processing M3___Experimental_Design

Start processing of:E1___Treatments

Table 4: E1___Treatments

	Variable	Value	Frequency
Treatment name.1	Treatment name	Early Bunch, 0kg N	1
Treatment name.2	Treatment name	Early Bunch, 60kg N	1
Treatment name.3	Treatment name	Early Bunch,120kg N	1
Treatment name.4	Treatment name	Early Bunch,240kg N	1
Treatment name.5	Treatment name	Florunner, 0kg N	1
Treatment name.6	Treatment name	Florunner, 60kg N	1

```
numeric_columns

 4  Variables      12  Observations
-------------------------------------------------------------------------------
Treatment.number
       n  missing distinct      Info     Mean  pMedian      Gmd      .05      .10      .25
      12        0       12         1      6.5      6.5    4.333     1.55     2.10     3.75
     .50      .75      .90      .95
    6.50     9.25    10.90    11.45

Value          1     2     3     4     5     6     7     8     9    10    11    12
Frequency      1     1     1     1     1     1     1     1     1     1     1     1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-------------------------------------------------------------------------------
Crop.ID
       n  missing distinct      Info     Mean  pMedian      Gmd
      12        0        3     0.895        2        2   0.9697

Value          1     2     3
Frequency      4     4     4
Proportion 0.333 0.333 0.333
-------------------------------------------------------------------------------
Fertilizer.schedule
       n  missing distinct      Info     Mean  pMedian      Gmd
      12        0        4     0.944      2.5      2.5    1.364

Value          1     2     3     4
Frequency      3     3     3     3
```
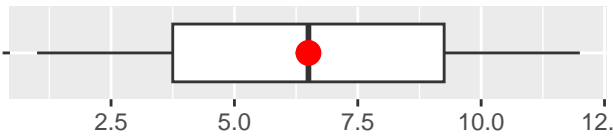
```
Proportion 0.25 0.25 0.25 0.25
-----------------------------------------------------------------------------------------
Harvest.schedule
       n  missing distinct     Info    Mean  pMedian     Gmd
      12       0        3     0.895       2        2   0.9697

Value          1     2     3
Frequency      4     4     4
Proportion 0.333 0.333 0.333
-----------------------------------------------------------------------------------------
```
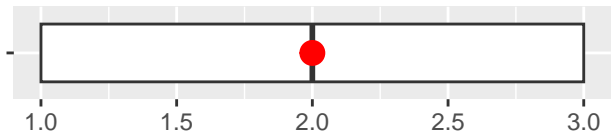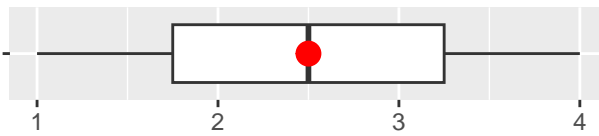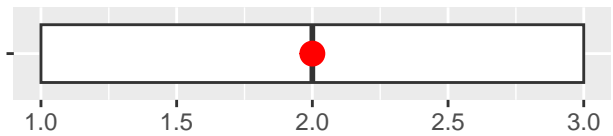
Treatment.number

Crop.ID

Fertilizer.schedule

Harvest.schedule

END Processing E1___Treatments

Start processing of:E2___Fields

Table 5: E2___Fields

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	1
Site	Site	UFGA	1
Field location	Field location	1	1
Soil ID	Soil ID	IBPN910015	1
Weather station ID	Weather station ID	UFGA	1

END Processing E2___Fields

Start processing of:E3___Plots

Table 6: E3___Plots

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	12
Site	Site	UFGA	12

```
numeric_columns

 2  Variables     12  Observations
-----------------------------------------------------------------------------------------
Plot.ID
       n  missing distinct     Info    Mean  pMedian     Gmd     .05     .10     .25
```

```
      12          0         12          1        6.5       6.5     4.333       1.55       2.10       3.75
     .50        .75        .90        .95
    6.50       9.25      10.90      11.45

Value          1      2      3      4      5      6      7      8      9     10     11     12
Frequency      1      1      1      1      1      1      1      1      1      1      1      1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
Treatment.number
       n  missing distinct      Info     Mean  pMedian       Gmd       .05       .10        .25
      12        0        12         1      6.5      6.5     4.333      1.55      2.10       3.75
     .50       .75       .90       .95
    6.50      9.25     10.90     11.45

Value          1      2      3      4      5      6      7      8      9     10     11     12
Frequency      1      1      1      1      1      1      1      1      1      1      1      1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
```


END Processing E3___Plots

Start processing of:E4___Crop_Information

Table 7: E4___Crop_Information

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	3
Site	Site	UFGA	3
Crop species	Crop species	Peanut	3
Cultivar.1	Cultivar	EARLY BUNCH	1
Cultivar.2	Cultivar	FLORUNNER, std	1
Cultivar.3	Cultivar	Non-Nodulated	1

END Processing E4___Crop_Information

Start processing of:E5___Planting

Table 8: E5___Planting

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	1
Site	Site	UFGA	1
Planting material	Planting material	dry seed	1
Planting distribution	Planting distribution	row	1

END Processing E5___Planting

Start processing of:E6___Irrigation

Table 9: E6___Irrigation

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	1
Site	Site	UFGA	1
Type of irrigation	Type of irrigation	sprinkle	1
Notes related to irrigation	Notes related to irrigation	Paper reports two irrigations, but original f	1

END Processing E6___Irrigation

Start processing of:E7___Fertilizer

Table 10: E7___Fertilizer

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	17
Site	Site	UFGA	17
Nutrient Source.1	Nutrient Source	Ammonium nitrate	9
Nutrient Source.2	Nutrient Source	compound fertilizer	4
Nutrient Source.3	Nutrient Source	gypsum	4
Placement.1	Placement	broadcast	8

```
numeric_columns

 6  Variables      17  Observations
--------------------------------------------------------------------------------
Fertilizer.schedule
      n  missing distinct     Info    Mean  pMedian     Gmd
     17        0       4    0.939   2.471     2.5   1.294


Value          1     2     3     4
Frequency      4     5     4     4
Proportion 0.235 0.294 0.235 0.235
--------------------------------------------------------------------------------
Date
       n   missing  distinct     Info     Mean   pMedian      Gmd
      17         0        4    0.909 1982-06-15     4545    38.29


Value    1982-05-10 1982-06-19 1982-07-15 1982-07-29
Frequency         7         3         4         3
Proportion    0.412     0.176     0.235     0.176
--------------------------------------------------------------------------------
Amount.of.elemental.N.applied
      n  missing distinct     Info    Mean  pMedian     Gmd
     17        0       4    0.882   24.71      20   32.65


Value          0    20    40    80
Frequency      8     3     3     3
Proportion 0.471 0.176 0.176 0.176
```

```
--------------------------------------------------------------------------------
Amount.of.elemental.P.applied
       n  missing distinct      Info      Mean
      17        0        2     0.542    0.5882

Value        0.0   2.5
Frequency     13     4
Proportion 0.765 0.235
--------------------------------------------------------------------------------
Amount.of.elemental.K.applied
       n  missing distinct      Info      Mean
      17        0        2     0.542     2.118

Value          0     9
Frequency     13     4
Proportion 0.765 0.235
--------------------------------------------------------------------------------
Depth.of.incorporation
       n  missing distinct      Info      Mean
      13        4        2     0.536     4.615

Value          0    20
Frequency     10     3
Proportion 0.769 0.231
--------------------------------------------------------------------------------
```
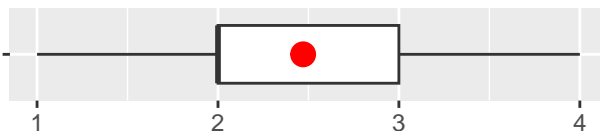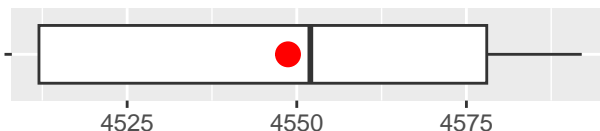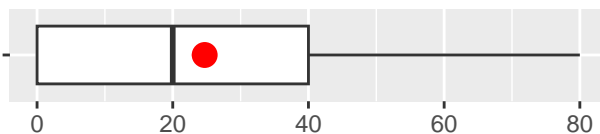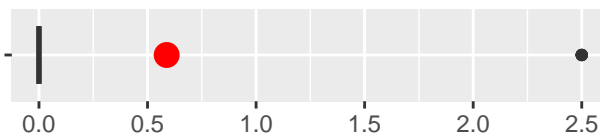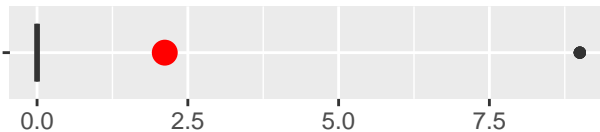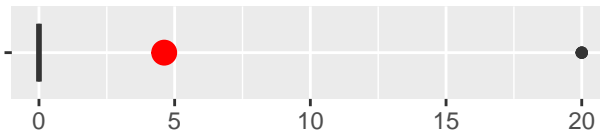

END Processing E7___Fertilizer

Start processing of:E9___Tillage

Table 11: E9___Tillage

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	3
Site	Site	UFGA	3

	Variable	Value	Frequency
Type of tillage operation.1	Type of tillage operation	broadcast fertilizer application	1
Type of tillage operation.2	Type of tillage operation	broadcast fertilizer application, gypsum	1
Type of tillage operation.3	Type of tillage operation	row planting with initial nitrogen	1
Notes related to operation.1	Notes related to operation	"Preplant" but date not given. No information	1

END Processing E9___Tillage

Start processing of:E8___Organic_Amendments

Table 12: E8___Organic_Amendments

Variable	Value	Frequency

END Processing E8___Organic_Amendments

Start processing of:E10___Chemical_Applications

Table 13: E10___Chemical_Applications

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	10
Site	Site	UFGA	10
Name of chemical applied.1	Name of chemical applied	Balan (N-buthyl-N-ethyl-a-a-a-trifluro-2,6-di	1
Name of chemical applied.2	Name of chemical applied	Bravo (tetrachloroisothalonitrile)	4
Name of chemical applied.3	Name of chemical applied	Sevin (1-naphyl N-methyl-carbamate)	4
Name of chemical applied.4	Name of chemical applied	Vernam (S-propyl dipropylthio-carbamate)	1

```
numeric_columns

 2  Variables     10  Observations
--------------------------------------------------------------------------------
Date
       n    missing   distinct      Info      Mean    pMedian       Gmd
      10          0          4     0.921 1982-07-07      4579     35.56


Value      1982-05-05 1982-07-09 1982-07-23 1982-08-06
Frequency           2          2          4          2
Proportion        0.2        0.2        0.4        0.2
--------------------------------------------------------------------------------
Chemicals.application.amount
       n   missing  distinct      Info      Mean
       2         8         2         1      1.82
```
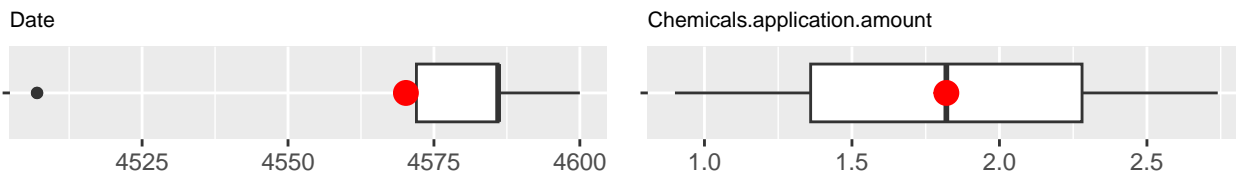
```
Value       0.90 2.74
Frequency      1    1
Proportion  0.5  0.5
-------------------------------------------------------------------------------------
```

Date

Chemicals.application.amount

```
        4525        4550        4575        4600      1.0      1.5      2.0      2.5
```

END Processing E10___Chemical_Applications

Start processing of:E11___Harvest

Table 14: E11___Harvest

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	3
Site	Site	UFGA	3
Crop species harvested	Crop species harvested	peanut	3
Harvest component	Harvest component	seed	3
Harvest method	Harvest method	hand	3

END Processing E11___Harvest

Start processing of:E12___Preplant_Soil

Table 15: E12___Preplant_Soil

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	9
Site	Site	UFGA	9

```
numeric_columns

 4  Variables     9  Observations
-------------------------------------------------------------------------------------
Depth.of.measurement..top.of.soil.layer
      n  missing distinct     Info     Mean  pMedian      Gmd
      9        0        9        1    57.22     52.5     62.5

Value         0     5    15    30    45    60    90   120   150
Frequency     1     1     1     1     1     1     1     1     1
Proportion 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111
-------------------------------------------------------------------------------------
Depth.of.measurement..bottom.of.soil.layer
      n  missing distinct     Info     Mean  pMedian      Gmd
      9        0        9        1    77.22       75    73.89

Value         5    15    30    45    60    90   120   150   180
```

```
Frequency        1     1     1     1     1     1     1     1     1
Proportion 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111
--------------------------------------------------------------------------------
Soil.water.content
       n  missing distinct     Info    Mean  pMedian      Gmd
       9        0        4    0.825  0.1078    0.086  0.04944

Value      0.076 0.086 0.130 0.258
Frequency      2     5     1     1
Proportion 0.222 0.556 0.111 0.111
--------------------------------------------------------------------------------
Ammonium.N
       n  missing distinct     Info    Mean  pMedian      Gmd
       9        0        3      0.8   1.067        1     0.55

Value        0.5   0.6   1.5
Frequency      3     1     5
Proportion 0.333 0.111 0.556
--------------------------------------------------------------------------------
```
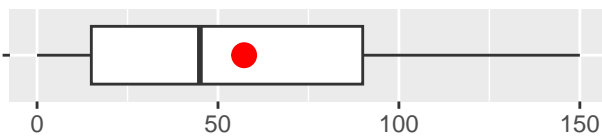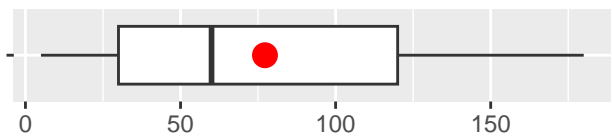

Depth.of.measurement..top.of.soil.layer

Depth.of.measurement..bottom.of.soil.layer

Soil.water.content
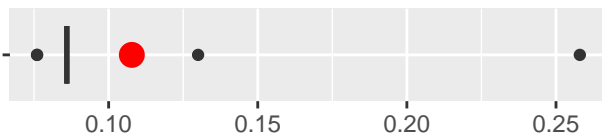
Ammonium.N
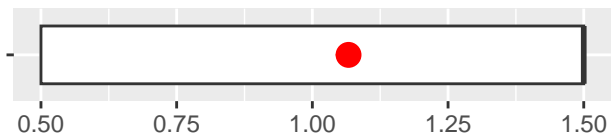
END Processing E12___Preplant_Soil

Start processing of:O1___Analysis_Methods

Table 16: O1___Analysis_Methods

Variable	Value	Frequency

END Processing O1___Analysis_Methods

Start processing of:O2___Yield_Summary

Table 17: O2___Yield_Summary

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	12
Site	Site	UFGA	12

```
numeric_columns
```

```
 14  Variables    12  Observations
--------------------------------------------------------------------------------
Plot.ID
      n  missing distinct     Info    Mean  pMedian     Gmd     .05     .10     .25
     12        0       12        1     6.5      6.5   4.333    1.55    2.10    3.75
    .50     .75      .90      .95
   6.50    9.25    10.90    11.45

Value           1     2     3     4     5     6     7     8     9    10    11    12
Frequency       1     1     1     1     1     1     1     1     1     1     1     1
Proportion  0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
Treatment.number
      n  missing distinct     Info    Mean  pMedian     Gmd     .05     .10     .25
     12        0       12        1     6.5      6.5   4.333    1.55    2.10    3.75
    .50     .75      .90      .95
   6.50    9.25    10.90    11.45

Value           1     2     3     4     5     6     7     8     9    10    11    12
Frequency       1     1     1     1     1     1     1     1     1     1     1     1
Proportion  0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
Seed.yield
      n  missing distinct     Info    Mean  pMedian     Gmd     .05     .10     .25
     12        0       11    0.997    1941     2090   639.5   884.6  1095.9  1811.2
    .50     .75      .90      .95
 2108.0  2250.8   2420.1   2561.6

Value         680  1052  1491  1918  2088  2108  2186  2245  2268  2437  2714
Frequency       1     1     1     1     1     2     1     1     1     1     1
Proportion  0.083 0.083 0.083 0.083 0.083 0.167 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
Single.seed.wt
      n  missing distinct     Info    Mean  pMedian     Gmd     .05     .10     .25
     12        0       12        1  0.6378    0.648  0.2721  0.3631  0.3733  0.4195
    .50     .75      .90      .95
 0.5820  0.9187   0.9432   0.9453

Value       0.356 0.369 0.412 0.422 0.571 0.575 0.589 0.620 0.913 0.936 0.944 0.947
Frequency       1     1     1     1     1     1     1     1     1     1     1     1
Proportion  0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
Seed.per.m2
      n  missing distinct     Info    Mean  pMedian     Gmd     .05     .10     .25
     12        0       12        1   317.8      316     107   208.6   223.6   238.8
    .50     .75      .90      .95
  321.5   383.2    433.5    445.6

Value         191   223   229   242   258   285   358   362   380   393   438   455
Frequency       1     1     1     1     1     1     1     1     1     1     1     1
Proportion  0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
LAI.max
      n  missing distinct     Info    Mean  pMedian     Gmd
```

```
      4         8         4         1      5.22      5.22      1.727

Value     3.93 4.30 5.75 6.90
Frequency     1    1    1    1
Proportion 0.25 0.25 0.25 0.25
--------------------------------------------------------------------------------
Tops.dry.wt
       n  missing distinct      Info      Mean   pMedian       Gmd       .05       .10       .25
      12        0       11     0.997     15350     14925      4612     10675     10970     11600
     .50      .75      .90       .95
   16050    18050    19550     20690

Value     10400 10900 11600 11900 14800 17300 17900 18000 18200 19700 21900
Frequency     1    1    2    1    1    1    1    1    1    1    1
Proportion 0.083 0.083 0.167 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
Pod.dry.wt
       n  missing distinct      Info      Mean   pMedian       Gmd       .05       .10       .25
      12        0       12         1      2664      2832     775.7      1333      1610      2570
     .50      .75      .90       .95
    2882     3024     3326      3394

Value      1068  1549  2160  2706  2830  2833  2932  2968  2984  3144  3346  3453
Frequency     1    1    1    1    1    1    1    1    1    1    1    1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
Harvest.index
       n  missing distinct      Info      Mean   pMedian       Gmd       .05       .10       .25
      12        0       12         1    0.1331    0.1335   0.06283   0.06695   0.07210   0.08650
     .50      .75      .90       .95
 0.12550  0.18550  0.20050   0.20280

Value     0.062 0.071 0.082 0.088 0.107 0.125 0.126 0.152 0.182 0.196 0.201 0.205
Frequency     1    1    1    1    1    1    1    1    1    1    1    1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
Pod.harvest.index
       n  missing distinct      Info      Mean   pMedian       Gmd       .05       .10       .25
      12        0       12         1    0.1837    0.1878   0.08373    0.1018    0.1064    0.1227
     .50      .75      .90       .95
  0.1605   0.2598   0.2800    0.2837

Value     0.098 0.105 0.119 0.124 0.149 0.157 0.164 0.193 0.256 0.271 0.281 0.287
Frequency     1    1    1    1    1    1    1    1    1    1    1    1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
--------------------------------------------------------------------------------
Threshing.percent
       n  missing distinct      Info      Mean   pMedian       Gmd       .05       .10       .25
      12        0       12         1     72.18     72.38     5.445     65.52     68.01     69.75
     .50      .75      .90       .95
   71.65    76.83    77.29     77.93

Value      62.6  67.9  69.0  70.0  70.8  71.1  72.2  72.7  76.7  77.2  77.3  78.7
Frequency     1    1    1    1    1    1    1    1    1    1    1    1
```

```
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-----------------------------------------------------------------------------------
Tops.nitrogen
       n  missing distinct      Info    Mean  pMedian      Gmd      .05      .10      .25
      12        0       12         1   214.7    216.9    68.11    116.6    135.8    191.0
     .50      .75      .90       .95
   215.7    259.3    264.3     281.2

Value       98.6 131.3 176.1 195.9 202.2 215.4 215.9 255.4 258.9 260.4 264.7 301.3
Frequency      1     1     1     1     1     1     1     1     1     1     1     1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-----------------------------------------------------------------------------------
Seed.nitrogen.tot
       n  missing distinct      Info    Mean  pMedian      Gmd      .05      .10      .25
      12        0       11     0.997   88.33    90.95    39.34    27.24    34.53    65.62
     .50      .75      .90       .95
  105.70   110.50   117.79    121.93

Value       20.2  33.0  48.3  71.4 100.4 103.9 107.5 109.7 110.5 118.6 126.0
Frequency      1     1     1     1     1     1     1     1     2     1     1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.167 0.083 0.083
-----------------------------------------------------------------------------------
Seed.nitrogen.conc.
       n  missing distinct      Info    Mean  pMedian      Gmd      .05      .10      .25
      12        0       12         1   4.353    4.335   0.9209    3.064    3.150    3.600
     .50      .75      .90       .95
   4.750    4.935    5.016     5.119

Value        2.97  3.14  3.24  3.72  4.64  4.74  4.76  4.87  4.92  4.98  5.02  5.24
Frequency       1     1     1     1     1     1     1     1     1     1     1     1
Proportion  0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-----------------------------------------------------------------------------------
```
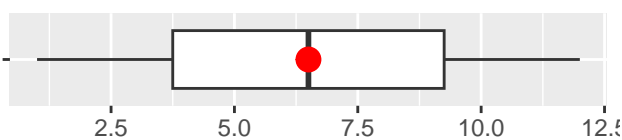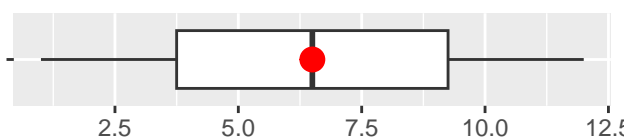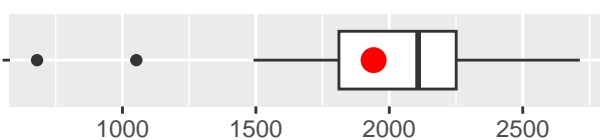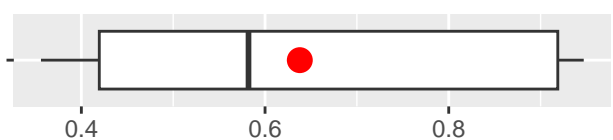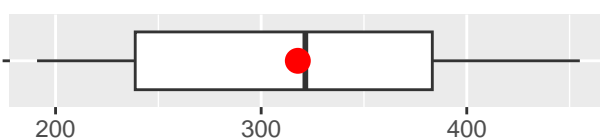
Boxplots: Tops.dry.wt, Pod.dry.wt, Harvest.index, Pod.harvest.index, Threshing.percent, Tops.nitrogen, Seed.nitrogen.tot, Seed.nitrogen.conc.
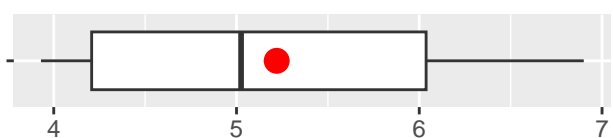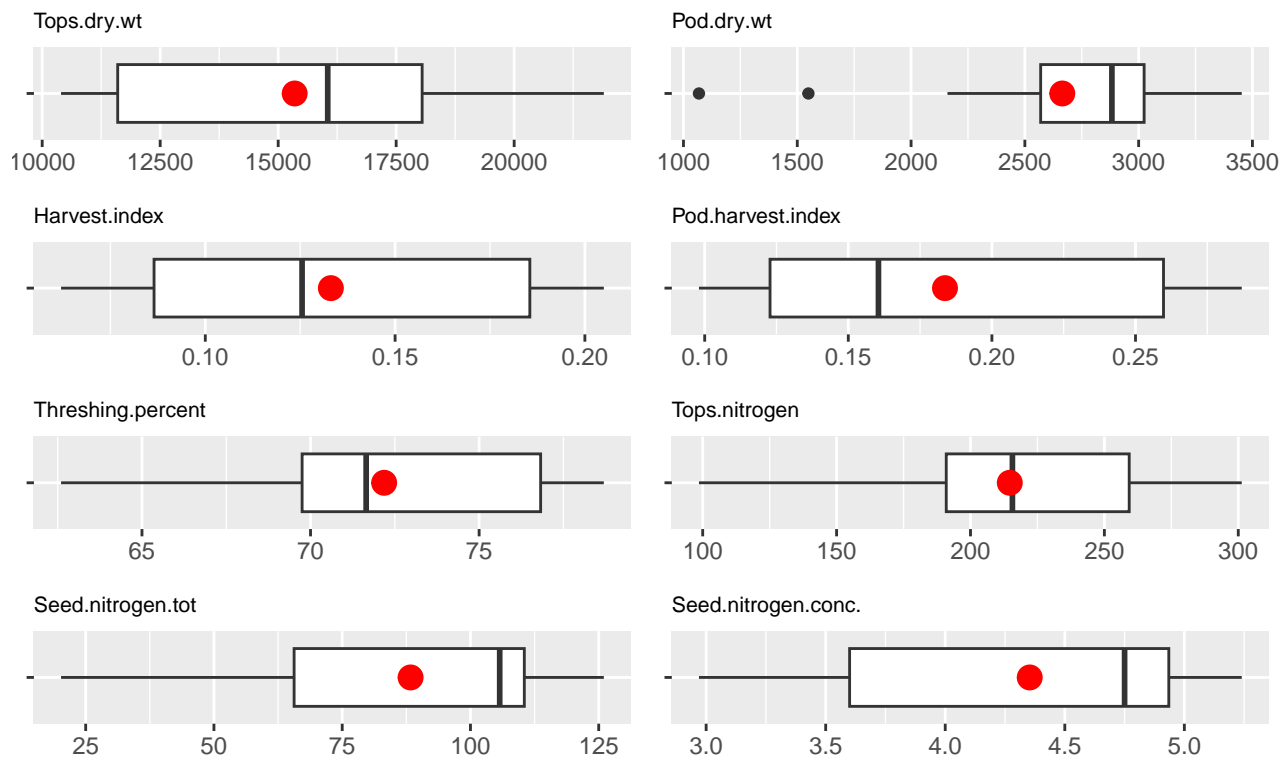
END Processing O2___Yield_Summary

Start processing of:O3___Crop_Growth

Table 18: O3___Crop_Growth

	Variable	Value	Frequency
Experiment ID	Experiment ID	UFGA8201	28
Site	Site	UFGA	28

```
numeric_columns

 18  Variables     28  Observations
----------------------------------------------------------------------------------
Treatment.number
       n  missing distinct      Info     Mean  pMedian       Gmd       .05       .10       .25
      28        0       12     0.978    4.214      3.5     3.423      1.00      1.00      2.00
     .50      .75      .90       .95
    3.00     5.25     9.30     10.65

Value          1     2     3     4     5     6     7     8     9    10    11    12
Frequency      5     5     5     5     1     1     1     1     1     1     1     1
Proportion 0.179 0.179 0.179 0.179 0.036 0.036 0.036 0.036 0.036 0.036 0.036 0.036
----------------------------------------------------------------------------------
Plot.ID
       n  missing distinct      Info     Mean  pMedian       Gmd       .05       .10       .25
      28        0       12     0.978    4.214      3.5     3.423      1.00      1.00      2.00
     .50      .75      .90       .95
    3.00     5.25     9.30     10.65
```

```
Value            1         2         3         4         5         6         7         8         9        10        11        12
Frequency        5         5         5         5         1         1         1         1         1         1         1         1
Proportion   0.179     0.179     0.179     0.179     0.036     0.036     0.036     0.036     0.036     0.036     0.036     0.036
-----------------------------------------------------------------------------------------------------------------
Sampling.date
        n    missing    distinct      Info       Mean     pMedian        Gmd
       28          0           7      0.981 1982-08-18        4612       41.06

Value    1982-06-29 1982-07-09 1982-08-03 1982-08-13 1982-09-12 1982-09-27 1982-10-02
Frequency         4          4          4          4          4          4          4
Proportion    0.143      0.143      0.143      0.143      0.143      0.143      0.143
-----------------------------------------------------------------------------------------------------------------
Leaf.area.index
        n    missing   distinct       Info       Mean    pMedian        Gmd        .05        .10        .25
       20          8         20          1      2.307       2.19      2.054     0.6100     0.6470     0.8025
      .50        .75        .90        .95
   1.6900     3.6725     4.4450     5.8075

Value      0.42 0.62 0.65 0.70 0.72 0.83 0.92 1.08 1.18 1.63 1.75 1.87 2.22 3.23 3.63 3.80
Frequency     1    1    1    1    1    1    1    1    1    1    1    1    1    1    1    1
Proportion 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05

Value      3.93 4.30 5.75 6.90
Frequency     1    1    1    1
Proportion 0.05 0.05 0.05 0.05
-----------------------------------------------------------------------------------------------------------------
Tops.dry.weight
        n    missing   distinct       Info       Mean    pMedian        Gmd        .05        .10        .25
       12         16         11      0.997      15350      14925       4612      10675      10970      11600
      .50        .75        .90        .95
    16050      18050      19550      20690

Value      10400 10900 11600 11900 14800 17300 17900 18000 18200 19700 21900
Frequency      1     1     2     1     1     1     1     1     1     1     1
Proportion 0.083 0.083 0.167 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-----------------------------------------------------------------------------------------------------------------
Pod.dry.weight
        n    missing   distinct       Info       Mean    pMedian        Gmd        .05        .10        .25
       12         16         12          1       2664       2832      775.7       1333       1610       2570
      .50        .75        .90        .95
     2882       3024       3326       3394

Value       1068  1549  2160  2706  2830  2833  2932  2968  2984  3144  3346  3453
Frequency      1     1     1     1     1     1     1     1     1     1     1     1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-----------------------------------------------------------------------------------------------------------------
Seed.dry.weight
        n    missing   distinct       Info       Mean    pMedian        Gmd        .05        .10        .25
       12         16         12          1       1948       2090        641      884.6     1095.9     1811.2
      .50        .75        .90        .95
   2147.0     2250.8     2420.1     2561.6

Value        680  1052  1491  1918  2088  2108  2186  2193  2245  2268  2437  2714
```

```
Frequency      1     1     1     1     1     1     1     1     1     1     1     1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-------------------------------------------------------------------------------------
Unit.seed.weight
       n  missing distinct     Info    Mean  pMedian     Gmd      .05      .10      .25
      12       16       12        1   637.8      648   272.1    363.1    373.3    419.5
     .50      .75      .90      .95
   582.0    918.8    943.2    945.4

Value        356   369   412   422   571   575   589   620   913   936   944   947
Frequency      1     1     1     1     1     1     1     1     1     1     1     1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-------------------------------------------------------------------------------------
Seed.number
       n  missing distinct     Info    Mean  pMedian     Gmd      .05      .10      .25
      12       16       12        1   317.8      316     107    208.6    223.6    238.8
     .50      .75      .90      .95
   321.5    383.2    433.5    445.6

Value        191   223   229   242   258   285   358   362   380   393   438   455
Frequency      1     1     1     1     1     1     1     1     1     1     1     1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-------------------------------------------------------------------------------------
Harvest.index
       n  missing distinct     Info    Mean  pMedian      Gmd       .05       .10       .25
      12       16       12        1  0.1331   0.1335  0.06283   0.06695   0.07210   0.08650
     .50      .75      .90      .95
 0.12550  0.18550  0.20050  0.20280

Value      0.062 0.071 0.082 0.088 0.107 0.125 0.126 0.152 0.182 0.196 0.201 0.205
Frequency      1     1     1     1     1     1     1     1     1     1     1     1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-------------------------------------------------------------------------------------
Pod.harvest.index
       n  missing distinct     Info    Mean  pMedian      Gmd       .05       .10       .25
      12       16       12        1  0.1837   0.1878  0.08373    0.1018    0.1064    0.1227
     .50      .75      .90      .95
  0.1605   0.2598   0.2800   0.2837

Value      0.098 0.105 0.119 0.124 0.149 0.157 0.164 0.193 0.256 0.271 0.281 0.287
Frequency      1     1     1     1     1     1     1     1     1     1     1     1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-------------------------------------------------------------------------------------
Shelling.percent
       n  missing distinct     Info    Mean  pMedian      Gmd      .05      .10      .25
      12       16       12        1   72.18    72.38    5.445    65.52    68.01    69.75
     .50      .75      .90      .95
   71.65    76.83    77.29    77.93

Value       62.6  67.9  69.0  70.0  70.8  71.1  72.2  72.7  76.7  77.2  77.3  78.7
Frequency      1     1     1     1     1     1     1     1     1     1     1     1
Proportion 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-------------------------------------------------------------------------------------
Tops.nitrogen
```

```
        n  missing distinct      Info     Mean  pMedian       Gmd      .05      .10      .25
       12       16       12         1    214.7    216.9     68.11    116.6    135.8    191.0
      .50      .75      .90      .95
    215.7    259.3    264.3    281.2

Value        98.6 131.3 176.1 195.9 202.2 215.4 215.9 255.4 258.9 260.4 264.7 301.3
Frequency       1     1     1     1     1     1     1     1     1     1     1     1
Proportion  0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-----------------------------------------------------------------------------------
Leaf.nitrogen
        n  missing distinct      Info     Mean  pMedian       Gmd      .05      .10      .25
       12       16       12         1    50.38    49.33     17.98    33.00    33.94    38.20
      .50      .75      .90      .95
    47.30    60.15    72.81    74.75

Value        32.4 33.5 37.9 38.3 39.1 43.0 51.6 56.3 58.9 63.9 73.8 75.9
Frequency       1     1     1     1     1     1     1     1     1     1     1     1
Proportion  0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-----------------------------------------------------------------------------------
Seed.nitrogen
        n  missing distinct      Info     Mean  pMedian       Gmd      .05      .10      .25
       12       16       11     0.997    88.33    90.95     39.34    27.24    34.53    65.62
      .50      .75      .90      .95
   105.70   110.50   117.79   121.93

Value        20.2 33.0 48.3 71.4 100.4 103.9 107.5 109.7 110.5 118.6 126.0
Frequency       1     1     1     1     1     1     1     1     2     1     1
Proportion  0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.167 0.083 0.083
-----------------------------------------------------------------------------------
Seed.N.concentration
        n  missing distinct      Info     Mean  pMedian       Gmd      .05      .10      .25
       12       16       12         1    4.341    4.338    0.8844    3.076    3.141    3.600
      .50      .75      .90      .95
    4.755    4.942    5.016    5.034

Value        3.01 3.13 3.24 3.72 4.64 4.74 4.77 4.86 4.93 4.98 5.02 5.05
Frequency       1     1     1     1     1     1     1     1     1     1     1     1
Proportion  0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
-----------------------------------------------------------------------------------
Leaf.N.concentration
        n  missing distinct      Info     Mean  pMedian       Gmd
       12       16        3     0.895    2.723     2.63    0.6788

Value        1.93  2.91  3.33
Frequency       4     4     4
Proportion  0.333 0.333 0.333
-----------------------------------------------------------------------------------
Stem.N.concentration
        n  missing distinct      Info     Mean  pMedian       Gmd
       12       16        3     0.895    1.047    1.015    0.2376

Value        0.77  1.11  1.26
Frequency       4     4     4
Proportion  0.333 0.333 0.333
```

Leaf.N.concentration

Stem.N.concentration
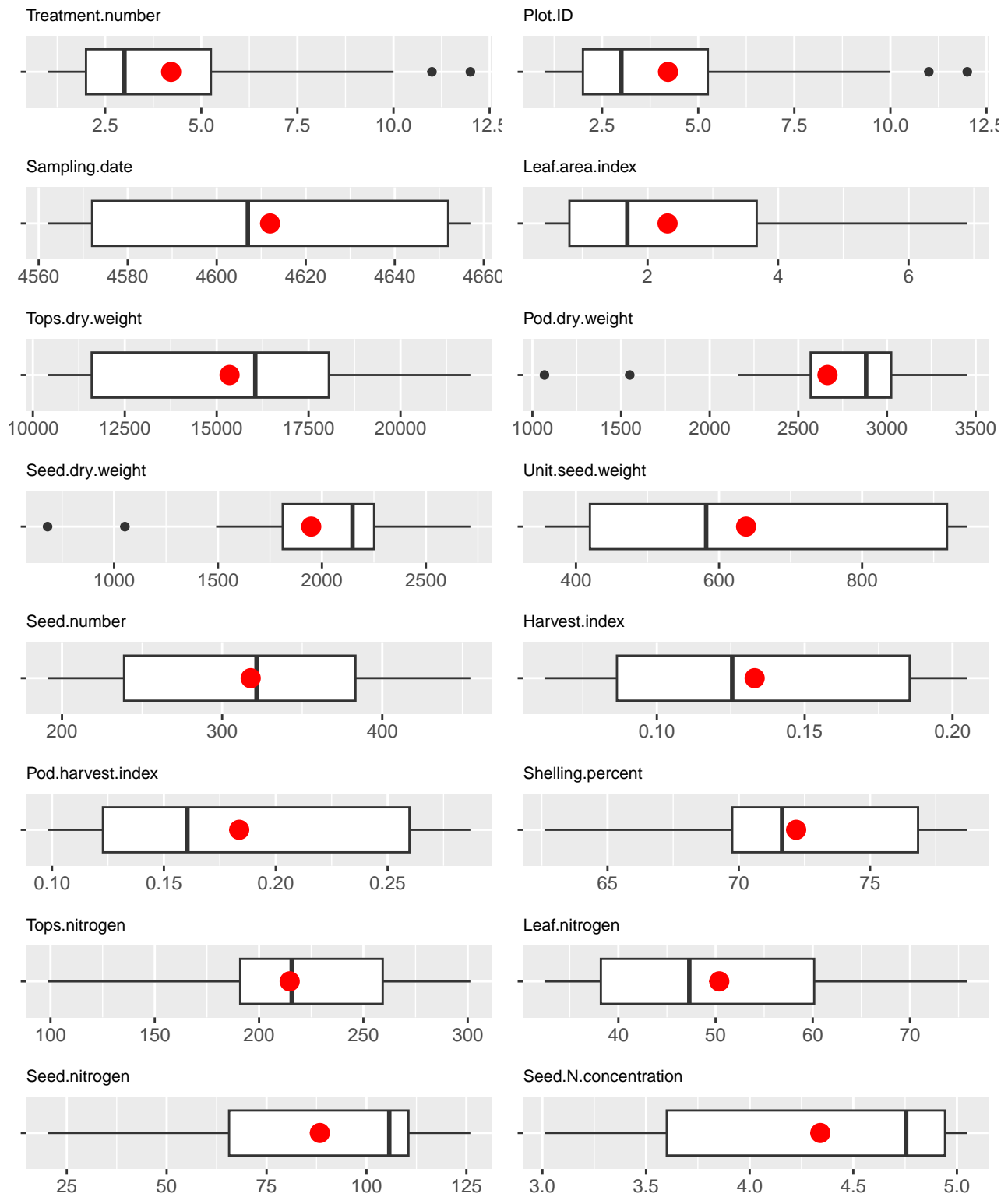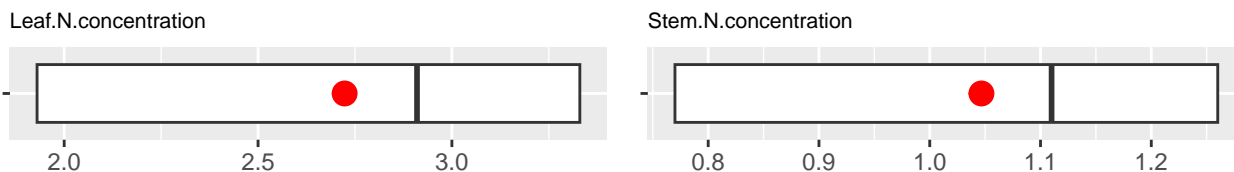
END Processing O3___Crop_Growth

Start processing of:O4___Crop_Health

Table 19: O4___Crop_Health

Variable	Value	Frequency

END Processing O4___Crop_Health

Start processing of:O5___Soil_Surface_Properties

Table 20: O5___Soil_Surface_Properties

Variable	Value	Frequency

END Processing O5___Soil_Surface_Properties

Start processing of:O6___Soil_Layer_Properties

Table 21: O6___Soil_Layer_Properties

Variable	Value	Frequency

END Processing O6___Soil_Layer_Properties

Start processing of:O7___Water

Table 22: O7___Water

Variable	Value	Frequency

END Processing O7___Water

Start processing of:S1___Soil_Metadata

Table 23: S1___Soil_Metadata

	Variable	Value	Frequency
Soil ID	Soil ID	IBPN910015	1
Soil name	Soil name	Millhopper Fine Sand	1
Soil classification	Soil classification	Loamy,silic,hyperth Gross. Paleudults	1
Soil classification system	Soil classification system	USDA	1

	Variable	Value	Frequency
Source of soil data	Source of soil data	DSSAT	1
Anonymize	Anonymize	N	1

END Processing S1___Soil_Metadata

Start processing of:S2___Soil_Layer_Properties

Table 24: S2___Soil_Layer_Properties

	Variable	Value	Frequency
Soil ID	Soil ID	IBPN910015	9

```
numeric_columns

 10  Variables      9  Observations
-------------------------------------------------------------------------------------------
Top.of.soil.layer
       n  missing distinct     Info     Mean  pMedian      Gmd
       9        0        9        1    57.22     52.5     62.5

Value          0     5    15    30    45    60    90   120   150
Frequency      1     1     1     1     1     1     1     1     1
Proportion 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111
-------------------------------------------------------------------------------------------
Bottom.of.soil.layer
       n  missing distinct     Info     Mean  pMedian      Gmd
       9        0        9        1    77.22       75    73.89

Value          5    15    30    45    60    90   120   150   180
Frequency      1     1     1     1     1     1     1     1     1
Proportion 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111 0.111
-------------------------------------------------------------------------------------------
Clay
       n  missing distinct     Info     Mean  pMedian      Gmd
       9        0        5    0.967    5.978      5.8    3.933

Value        0.9   4.6   5.8   8.3   9.6
Frequency      2     1     2     2     2
Proportion 0.222 0.111 0.222 0.222 0.222
-------------------------------------------------------------------------------------------
Silt
       n  missing distinct     Info     Mean  pMedian      Gmd
       9        0        5    0.967    6.267      5.4      3.5

Value        3.6   4.2   5.4   6.4  11.8
Frequency      2     2     2     1     2
Proportion 0.222 0.222 0.222 0.111 0.222
-------------------------------------------------------------------------------------------
Sand
       n  missing distinct     Info     Mean  pMedian      Gmd
       9        0        5    0.967    87.76     87.7    1.267
```

```
Value         86.2  87.3  88.1  88.8  89.0
Frequency        2     2     2     2     1
Proportion   0.222 0.222 0.222 0.222 0.111
--------------------------------------------------------------------------------
Organic.matter
       n  missing distinct      Info     Mean  pMedian      Gmd
       9        0        6     0.958   0.2722      0.2   0.3372

Value        0.03  0.09  0.20  0.28  0.69  0.90
Frequency       3     1     2     1     1     1
Proportion  0.333 0.111 0.222 0.111 0.111 0.111
--------------------------------------------------------------------------------
Bulk.density
       n  missing distinct      Info     Mean  pMedian      Gmd
       9        0        8     0.992    1.491    1.465     0.13

Value        1.36  1.40  1.43  1.46  1.47  1.48  1.57  1.79
Frequency       1     1     1     2     1     1     1     1
Proportion  0.111 0.111 0.111 0.222 0.111 0.111 0.111 0.111
--------------------------------------------------------------------------------
Wilting.point
       n  missing distinct      Info     Mean  pMedian      Gmd
       9        0        5     0.833  0.02811    0.023  0.01211

Value       0.020 0.021 0.023 0.027 0.070
Frequency       1     1     5     1     1
Proportion  0.111 0.111 0.556 0.111 0.111
--------------------------------------------------------------------------------
Field.capacity
       n  missing distinct      Info     Mean  pMedian      Gmd
       9        0        4     0.825   0.1078    0.086  0.04944

Value       0.076 0.086 0.130 0.258
Frequency       2     5     1     1
Proportion  0.222 0.556 0.111 0.111
--------------------------------------------------------------------------------
Saturated.hydraulic.conductivity
       n  missing distinct      Info     Mean  pMedian      Gmd
       9        0        7     0.983    14.68    14.15    13.62

Value         0.1   0.3   7.4  15.8  17.5  27.6  28.0
Frequency       1     1     2     1     1     1     2
Proportion  0.111 0.111 0.222 0.111 0.111 0.111 0.222
--------------------------------------------------------------------------------
```


Top.of.soil.layer

Bottom.of.soil.layer

Clay

2.5 5.0 7.5 10.0

Silt

5.0 7.5 10.0

Sand

87 88 89

Organic.matter

0.00 0.25 0.50 0.75

Bulk.density

1.4 1.5 1.6 1.7 1.8

Wilting.point

0.02 0.03 0.04 0.05 0.06 0.07

Field.capacity

0.10 0.15 0.20 0.25

Saturated.hydraulic.conductivity

0 10 20
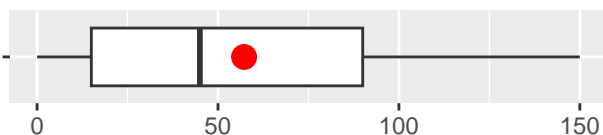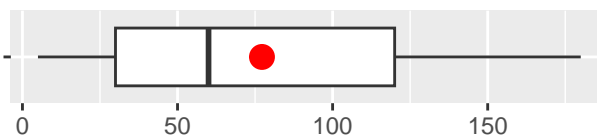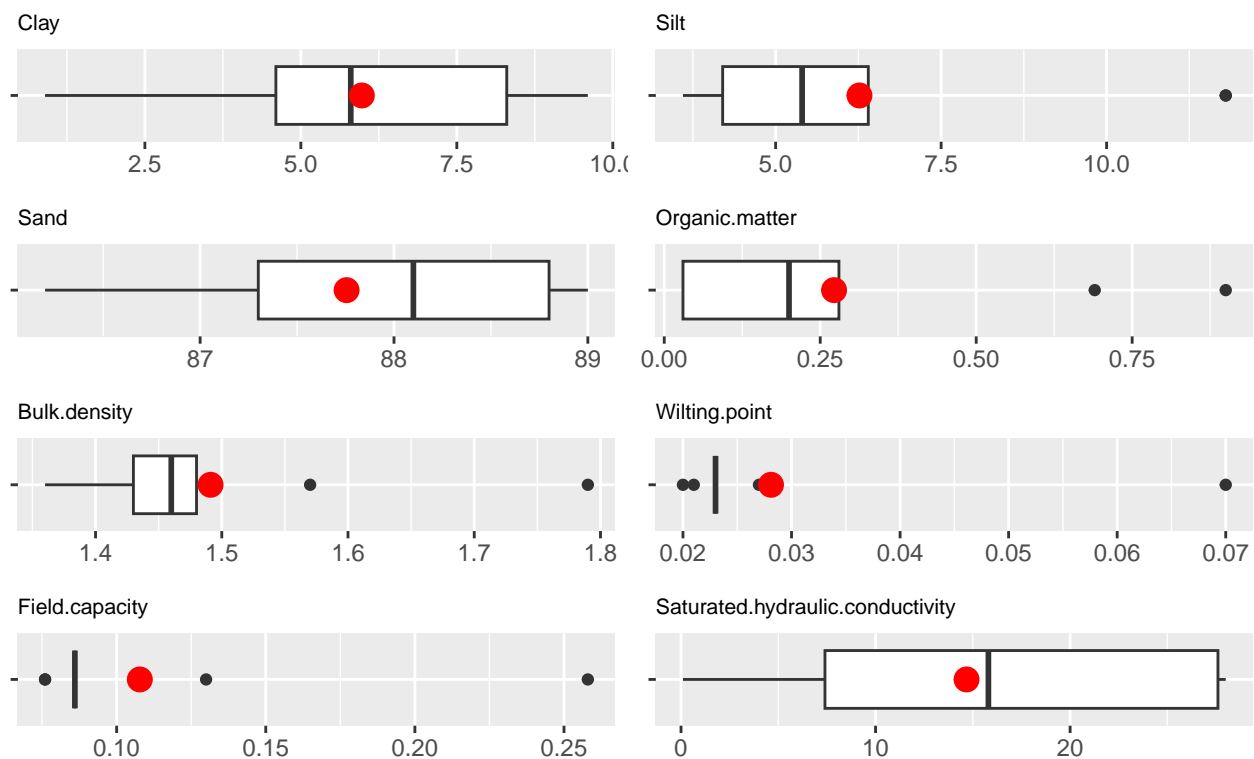
END Processing S2___Soil_Layer_Properties

Start processing of:W1___Weather_Station_Metadata

Table 25: W1___Weather_Station_Metadata

	Variable	Value	Frequency
Weather station ID	Weather station ID	UFGA	1
Weather station name	Weather station name	Gainesville,Florida,USA	1
Anonymize	Anonymize	N	1

END Processing W1___Weather_Station_Metadata

Start processing of:W2___Daily_Weather_Data

Table 26: W2___Daily_Weather_Data

	Variable	Value	Frequency
Weather station ID	Weather station ID	UFGA	365

```
numeric_columns

 6  Variables      365  Observations
--------------------------------------------------------------------------------
Date
        n    missing   distinct        Info       Mean     pMedian        Gmd         .05
      365          0        365           1 1982-07-02       4565         122 1982-01-19
      .10        .25        .50         .75        .90         .95
```

23

```
1982-02-06 1982-04-02 1982-07-02 1982-10-01 1982-11-24 1982-12-12

lowest : 1982-01-01 1982-01-02 1982-01-03 1982-01-04 1982-01-05
highest: 1982-12-27 1982-12-28 1982-12-29 1982-12-30 1982-12-31
-------------------------------------------------------------------------------
Minimum.daily.air.temperature
       n  missing distinct     Info     Mean  pMedian      Gmd      .05      .10      .25
     365        0       49    0.998    15.69     16.4    6.778     2.32     6.34    12.80
     .50      .75      .90      .95
   16.70    21.10    22.20    22.80

lowest : -7.8 -6.7 -3.3 -1.7 -1.1, highest: 22.2 22.8 23.3 23.9 26.1
-------------------------------------------------------------------------------
Maximum.daily.air.temperature
       n  missing distinct     Info     Mean  pMedian      Gmd      .05      .10      .25
     365        0       40    0.998    28.27    28.65    5.232    18.90    21.34    25.60
     .50      .75      .90      .95
   28.90    32.20    33.30    34.30

lowest : 10   12.2 12.8 14.4 15.6, highest: 33.3 33.9 34.4 35   35.6
-------------------------------------------------------------------------------
Daily.precipitation
       n  missing distinct     Info     Mean  pMedian      Gmd      .05      .10      .25
     365        0       67     0.69    4.232      0.4    7.448     0.00     0.00     0.00
     .50      .75      .90      .95
    0.00     1.30    14.00    24.06

lowest : 0    0.3  0.5  0.8  1   , highest: 64.5 67.3 70.1 81   98.8
-------------------------------------------------------------------------------
Solar.radiation
       n  missing distinct     Info     Mean  pMedian      Gmd      .05      .10      .25
     365        0      156        1    14.65     14.6    6.967      5.0      7.0     10.3
     .50      .75      .90      .95
    14.0     19.0     23.5     24.6

lowest : 0.8  2    3    3.1  3.3 , highest: 25.9 26.3 26.5 26.7 26.9
-------------------------------------------------------------------------------
PAR
       n  missing distinct     Info     Mean  pMedian      Gmd      .05      .10      .25
     365        0      227        1    29.66    29.45    13.63    10.36    14.44    22.00
     .50      .75      .90      .95
   28.00    37.60    46.36    50.94

lowest : 2.2  5    6    6.4  6.6 , highest: 55.7 56.1 56.2 56.3 57.2
-------------------------------------------------------------------------------
```


24
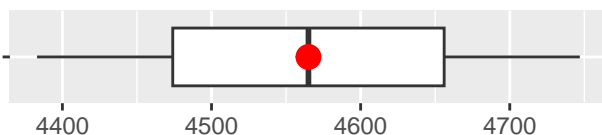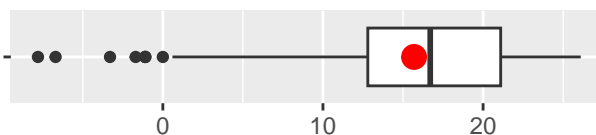
Maximum.daily.air.temperature
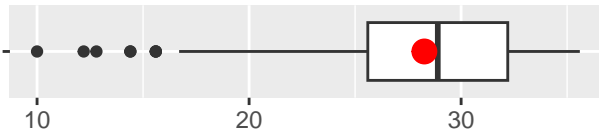
Daily.precipitation
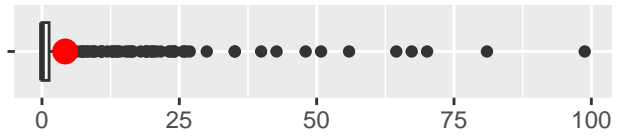
Solar.radiation
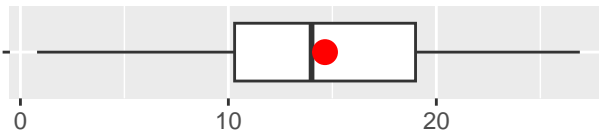
PAR
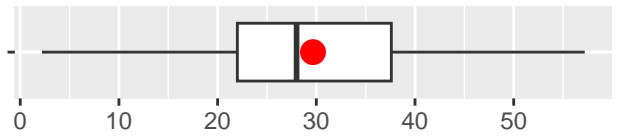
END Processing W2___Daily_Weather_Data

2.2. Correct Dates? Formats correct? Events sequenced as expected?

Dates of key management events such as plantings, irrigations and harvests are sometimes entered incorrectly. A common problem is inversion of days and months (is '3/5/2021' "March 5, 2021" or "3 may 2021"?). Dates also may be entered in non-standard formats such as text. We prefer that the internal data format be used and data be expressed in ISO format (e.g., 2023-03-13 or yyyy-mm-dd). The various workbook formats for expressing dates are still acceptable so long as the data are truly recorded as dates.

2.2.1 Checking that date variables are formatted as dates

We first test whether dates reported for management, measurements and weather are in an acceptable date format. Not all date variables may be tested.

```
[1] "The variable 'Planting date' in 'E5__Planting' is valid."

[1] "The variable 'Date of irrigation' in 'E6__Irrigation' is valid."

[1] "The variable 'Date' in 'E7__Fertilizer' is valid."

[1] "No rows with data in 'E8__Organic_Amendments'. Skipping date test."

[1] "The variable 'Date' in 'E9__Tillage' is valid."

[1] "The variable 'Date' in 'E10__Chemical_Applications' is valid."

[1] "The variable 'Harvest date' in 'E11__Harvest' is given as numeric and cannot be processed as given

[1] "The variable 'Sampling date' in 'E12__Preplant_Soil' is valid."

[1] "The variable 'Sampling date' in 'O3__Crop_Growth' is valid."

[1] "No rows with data in 'O4__Crop_Health'. Skipping date test."

[1] "No rows with data in 'O5__Soil_Surface_Properties'. Skipping date test."

[1] "No rows with data in 'O6__Soil_Layer_Properties'. Skipping date test."

[1] "The variable 'Date' in 'W2__Daily_Weather_Data' is valid."
```

2.2.2 Checking dates for management events using a timeline

As a further check of key dates reported for crop management, we plot management events for each combination of Experiment, Site and Year along a timeline. To reduce the potential number of plots, data from different treatments and replicates are pooled together. This means some timelines may include multiple instances of plantings, fertilizer applications, harvests or other events. We currently do not consider crop phenology such as flowering or maturity dates.

Timelines for Experiments, Sites and Years
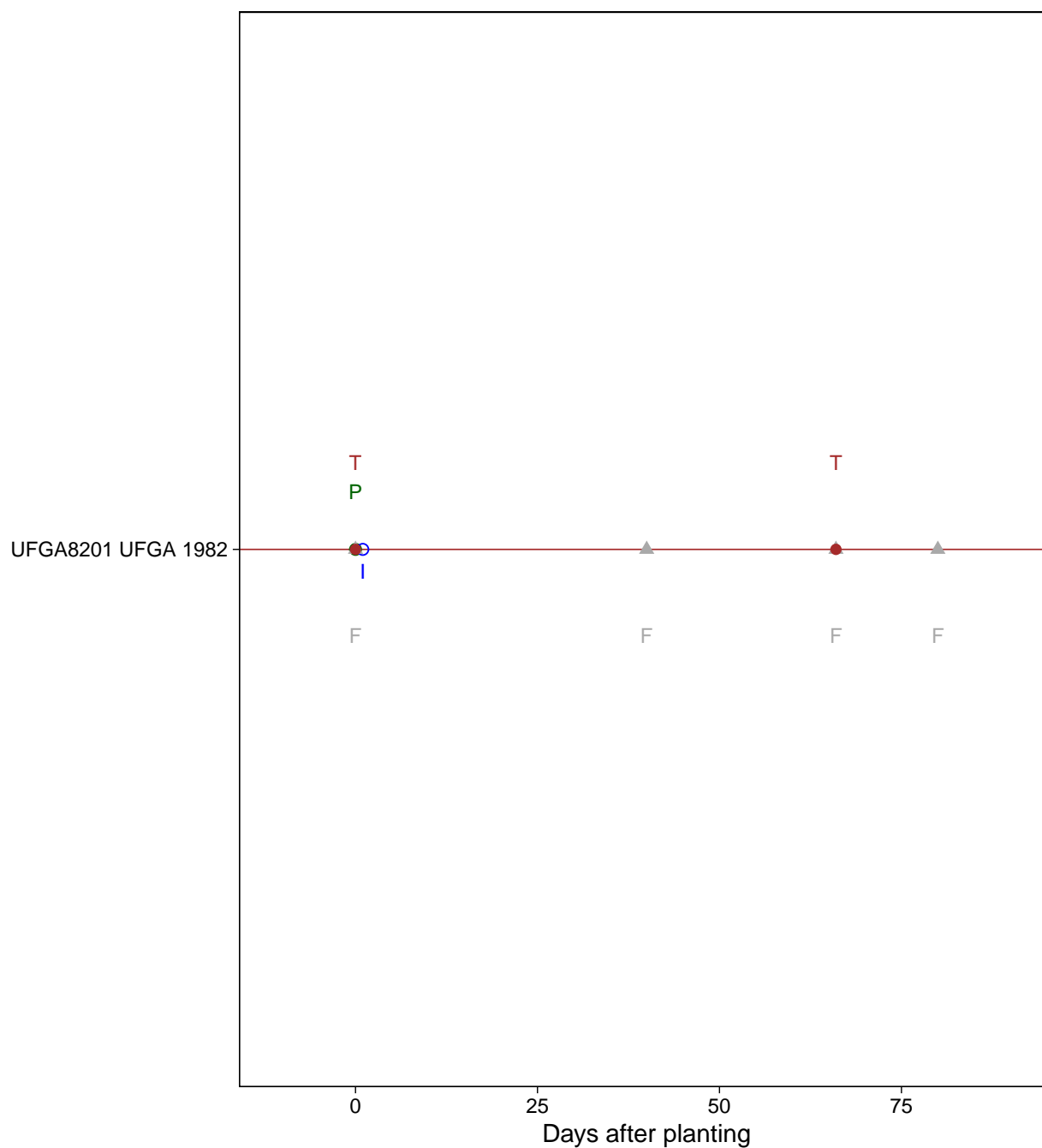
Replicates or treatments having identical event dates are shown as a single line.

2.3. Correct geocoordinates? Are locations mapped as expected?

Experience shows that datasets often have errors in location data. This section checks that any reported geocoordinates are roughly correct by mapping. Geocoordinates may appear in four sheets:

- M2. Sites
- E2. Fields
- S1. Soil Metadata
- W1. Weather Station Metadata

To facilitate processing, we extract the geocoordinates and the location name, and add as 'Source' the name of the individual sheet containing the data.

2.3.1. List of all expected geocoordinates

Source	Location	Lat	Long
M2. Sites	UFGA	29.64	-82.37
E2. Fields	UFGA	NA	NA
S1. Soil Meta	IBPN910015	29.63	-82.37
W1. Weather	Gainesville,Florida,USA	29.63	-82.37

2.3.2. Displaying the reference map of Florida with any reported locations

Here we use a map of Florida as the base. If latitude or longitude values are very far off (e.g., if the values are reversed or longitude is assigned a positive value for anywhere in the Americas), the map will display, but it may be distorted and not look like the expected base map of Florida.

Geocoordinates of Reported Sites for Research, Soils or Weather
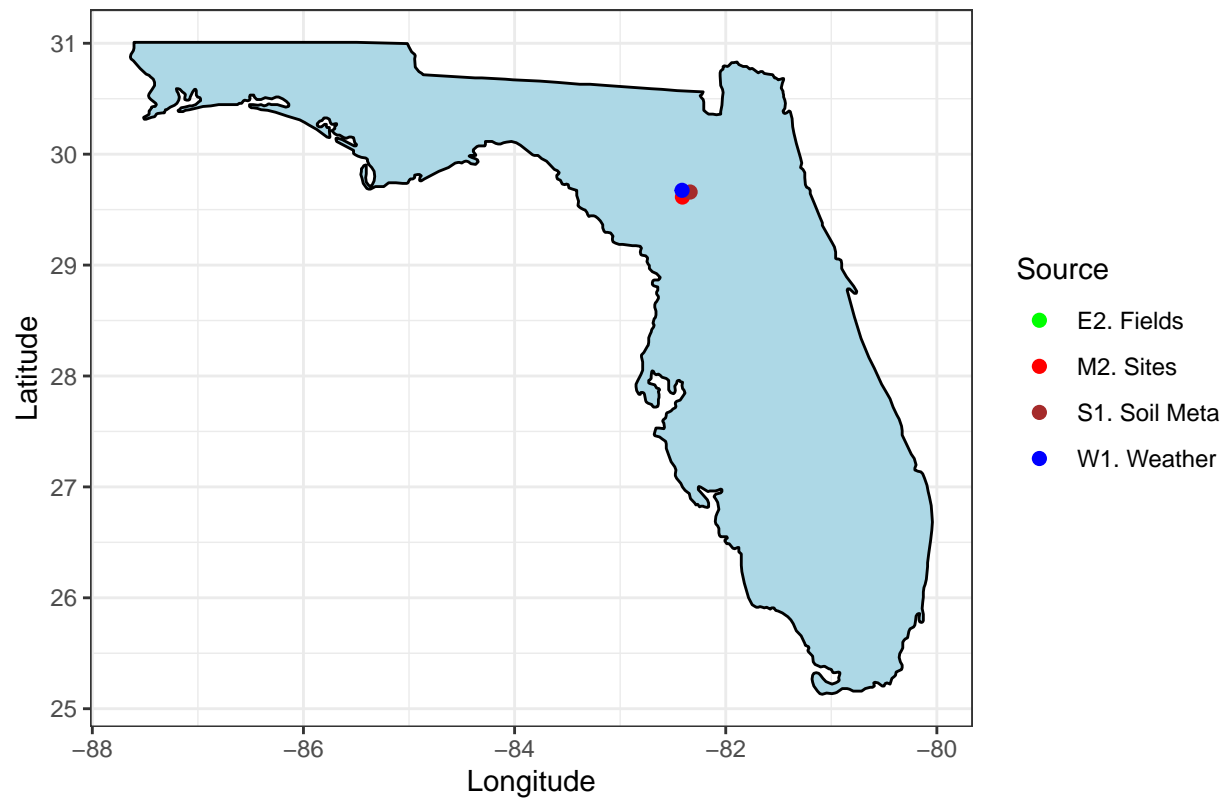
The option position = position_jitter is used so that if there are a large number of points with nearly identical locations, these are spread out slightly.

2.4. Completeness of sheets: Checking whether sheets present in the template are missing from the dataset

Users may add sheets as needed but are discouraged from deleting sheets.

```
The sheet names match.
```

2.5. Completeness of data in individual sheets

To assess completeness, we need to know how many variables actually have data values (e.g., are not empty cells). Below is a count of total values for variables used in each sheet. To avoid the output being split into two sections, variable names that are longer than 30 characters are truncated.

```
Total Non-Missing Values across all sheets: 4286 \n
```

```
Total Missing Values across all sheets: 811 \n
```

Sheet_name	Variable	Non_NA	Missing
M1. Experiments	Experiment name	1	0
M1. Experiments	Experiment ID	1	0
M1. Experiments	Research data owner	1	0
M1. Experiments	Institutional data owner	1	0
M1. Experiments	Contributor e-mail	0	1
M1. Experiments	Publication journal & volume	1	0
M1. Experiments	Link to document	1	0
M1. Experiments	Publication DOI	1	0
M1. Experiments	Anonymize	1	0
M1. Experiments	Data release date	0	1
M2. Sites	Site	1	0
M2. Sites	Local name for experiment site	1	0
M2. Sites	State	1	0
M2. Sites	County	1	0
M2. Sites	Town or other	1	0
M2. Sites	Latitude	1	0
M2. Sites	Longitude	1	0
M3. Experimental Design	Experiment ID	1	0
M3. Experimental Design	Site	1	0
M3. Experimental Design	Rate treatments	1	0
M3. Experimental Design	Replicates	1	0
M3. Experimental Design	Treatment structure	1	0
M3. Experimental Design	Type of experiment	1	0
M3. Experimental Design	Main effect 1	1	0
M3. Experimental Design	Main effect 2	1	0
M3. Experimental Design	Plot width	0	1
M3. Experimental Design	Plot length	1	0
E1. Treatments	Treatment number	12	0
E1. Treatments	Treatment name	12	0
E1. Treatments	Experiment ID	12	0
E1. Treatments	Site	12	0
E1. Treatments	Field location	12	0

E1. Treatments	Study year	12	0
E1. Treatments	Crop ID	12	0
E1. Treatments	Planting schedule	12	0
E1. Treatments	Irrigation schedule	12	0
E1. Treatments	Fertilizer schedule	12	0
E1. Treatments	Organic amendments schedule	12	0
E1. Treatments	Chemical applications schedule	12	0
E1. Treatments	Tillage schedule	12	0
E1. Treatments	Harvest schedule	12	0
E1. Treatments	Soil initial conditions ID	12	0
E1. Treatments	Comments about treatment	0	12
E2. Fields	Experiment ID	1	0
E2. Fields	Site	1	0
E2. Fields	Field location	1	0
E2. Fields	Latitude	0	1
E2. Fields	Longitude	0	1
E2. Fields	Soil ID	1	0
E2. Fields	Weather station ID	1	0
E2. Fields	Distance to weather station	0	1
E2. Fields	Field area	0	1
E2. Fields	Field length to width ratio	0	1
E2. Fields	Field slope	0	1
E2. Fields	Drainage type	0	1
E2. Fields	Water table depth	0	1
E2. Fields	Type of organic matter	0	1
E2. Fields	Dry mass of surface organic matter	1	0
E2. Fields	Nitrogen concentration in surface organic matter	1	0
E2. Fields	Phosphorus concentration in surface organic matter	1	0
E2. Fields	Portion of residue incorporated	1	0
E2. Fields	Depth of residue incorporation	1	0
E3. Plots	Plot ID	12	0
E3. Plots	Experiment ID	12	0
E3. Plots	Site	12	0
E3. Plots	Field location	12	0
E3. Plots	Treatment number	12	0
E3. Plots	Replicate	12	0
E4. Crop Information	Experiment ID	3	0
E4. Crop Information	Site	3	0
E4. Crop Information	Year	3	0
E4. Crop Information	Crop ID	3	0
E4. Crop Information	Crop species	3	0
E4. Crop Information	Cultivar	3	0
E4. Crop Information	Intended crop usage	3	0
E4. Crop Information	Cultivar notes	1	2
E5. Planting	Experiment ID	1	0
E5. Planting	Site	1	0
E5. Planting	Year	1	0
E5. Planting	Planting schedule	1	0
E5. Planting	Planting date	1	0
E5. Planting	Row spacing	1	0
E5. Planting	Planting density	1	0

E5. Planting	Plant density at emergence	0	1
E5. Planting	Planting material	1	0
E5. Planting	Planting distribution	1	0
E6. Irrigation	Experiment ID	1	0
E6. Irrigation	Site	1	0
E6. Irrigation	Year	1	0
E6. Irrigation	Irrigation schedule	1	0
E6. Irrigation	Date of irrigation	1	0
E6. Irrigation	Type of irrigation	1	0
E6. Irrigation	Amount of irrigation	1	0
E6. Irrigation	Notes related to irrigation	1	0
E7. Fertilizer	Experiment ID	17	0
E7. Fertilizer	Site	17	0
E7. Fertilizer	Year	17	0
E7. Fertilizer	Fertilizer schedule	17	0
E7. Fertilizer	Date	17	0
E7. Fertilizer	Nutrient Source	17	0
E7. Fertilizer	Amount of elemental N applied	17	0
E7. Fertilizer	Amount of elemental P applied	17	0
E7. Fertilizer	Amount of elemental K applied	17	0
E7. Fertilizer	Placement	17	0
E7. Fertilizer	Depth of incorporation	13	4
E7. Fertilizer	Analysis	13	4
E7. Fertilizer	Application timing	17	0
E7. Fertilizer	Notes related to application	8	9
E9. Tillage	Experiment ID	3	0
E9. Tillage	Site	3	0
E9. Tillage	Year	3	0
E9. Tillage	Tillage schedule	3	0
E9. Tillage	Date	3	0
E9. Tillage	Type of tillage operation	3	0
E9. Tillage	Depth of incorporation	3	0
E9. Tillage	Notes related to operation	3	0
E8. Organic Amendments	Experiment ID	0	0
E8. Organic Amendments	Site	0	0
E8. Organic Amendments	Year	0	0
E8. Organic Amendments	Organic amendments schedule	0	0
E8. Organic Amendments	Date	0	0
E8. Organic Amendments	Type of organic matter	0	0
E8. Organic Amendments	Amount of organic matter	0	0
E8. Organic Amendments	Placement	0	0
E8. Organic Amendments	Depth of incorporation	0	0
E8. Organic Amendments	N concentration	0	0
E8. Organic Amendments	Notes related to application	0	0
E10. Chemical Applications	Experiment ID	10	0
E10. Chemical Applications	Site	10	0
E10. Chemical Applications	Year	10	0
E10. Chemical Applications	Chemical application schedule	10	0
E10. Chemical Applications	Date	10	0
E10. Chemical Applications	Name of chemical applied	10	0
E10. Chemical Applications	Chemicals application amount	10	0

E10. Chemical Applications	Chemicals application method	10	0
E10. Chemical Applications	Depth of application	10	0
E10. Chemical Applications	Chemicals application target	0	10
E10. Chemical Applications	Notes related to application	10	0
E11. Harvest	Experiment ID	3	0
E11. Harvest	Site	3	0
E11. Harvest	Year	3	0
E11. Harvest	Harvest schedule	3	0
E11. Harvest	Harvest date	0	3
E11. Harvest	Crop species harvested	3	0
E11. Harvest	Harvest component	3	0
E11. Harvest	Harvest method	3	0
E11. Harvest	Main product harvested	3	0
E11. Harvest	By-product harvested	3	0
E12. Preplant Soil	Experiment ID	9	0
E12. Preplant Soil	Site	9	0
E12. Preplant Soil	Year	9	0
E12. Preplant Soil	Soil initial conditions ID	9	0
E12. Preplant Soil	Sampling date	9	0
E12. Preplant Soil	Depth of measurement, top of soil layer	9	0
E12. Preplant Soil	Depth of measurement, bottom of soil layer	9	0
E12. Preplant Soil	Soil water content	9	0
E12. Preplant Soil	Nitrate N	9	0
E12. Preplant Soil	Ammonium N	9	0
E12. Preplant Soil	Stable organic C	0	9
O1. Analysis Methods	Experiment ID	0	0
O1. Analysis Methods	Full parameter name	0	0
O1. Analysis Methods	Header name (in data file)	0	0
O1. Analysis Methods	Unit	0	0
O1. Analysis Methods	Matrix	0	0
O1. Analysis Methods	Analytical laboratory	0	0
O1. Analysis Methods	Analysis method	0	0
O1. Analysis Methods	EPA method	0	0
O1. Analysis Methods	Computation method	0	0
O2. Yield Summary	Experiment ID	12	0
O2. Yield Summary	Site	12	0
O2. Yield Summary	Year	12	0
O2. Yield Summary	Plot ID	12	0
O2. Yield Summary	Treatment number	12	0
O2. Yield Summary	Replicate	0	12
O2. Yield Summary	Seed yield	12	0
O2. Yield Summary	Single seed wt	12	0
O2. Yield Summary	Seed per m2	12	0
O2. Yield Summary	LAI max	12	0
O2. Yield Summary	Tops dry wt	12	0
O2. Yield Summary	Pod dry wt	12	0
O2. Yield Summary	Harvest index	12	0
O2. Yield Summary	Pod harvest index	12	0
O2. Yield Summary	Threshing percent	12	0
O2. Yield Summary	Tops nitrogen	12	0
O2. Yield Summary	Seed nitrogen tot	12	0

O2. Yield Summary	Seed nitrogen conc.	12	0
O3. Crop Growth	Experiment ID	28	0
O3. Crop Growth	Site	28	0
O3. Crop Growth	Year	28	0
O3. Crop Growth	Treatment number	28	0
O3. Crop Growth	Replicate	28	0
O3. Crop Growth	Plot ID	28	0
O3. Crop Growth	Sampling date	28	0
O3. Crop Growth	Leaf area index	28	0
O3. Crop Growth	Tops dry weight	28	0
O3. Crop Growth	Pod dry weight	28	0
O3. Crop Growth	Seed dry weight	28	0
O3. Crop Growth	Unit seed weight	28	0
O3. Crop Growth	Seed number	28	0
O3. Crop Growth	Harvest index	28	0
O3. Crop Growth	Pod harvest index	28	0
O3. Crop Growth	Shelling percent	28	0
O3. Crop Growth	Tops nitrogen	28	0
O3. Crop Growth	Leaf nitrogen	28	0
O3. Crop Growth	Seed nitrogen	28	0
O3. Crop Growth	Seed N concentration	28	0
O3. Crop Growth	Leaf N concentration	28	0
O3. Crop Growth	Stem N concentration	28	0
O4. Crop Health	Experiment ID	0	0
O4. Crop Health	Site	0	0
O4. Crop Health	Year	0	0
O4. Crop Health	Treatment number	0	0
O4. Crop Health	Replicate	0	0
O4. Crop Health	Plot ID	0	0
O4. Crop Health	Sampling date	0	0
O4. Crop Health	Field notes	0	0
O5. Soil Surface Properties	Experiment ID	0	0
O5. Soil Surface Properties	Site	0	0
O5. Soil Surface Properties	Year	0	0
O5. Soil Surface Properties	Treatment number	0	0
O5. Soil Surface Properties	Replicate	0	0
O5. Soil Surface Properties	Plot ID	0	0
O5. Soil Surface Properties	Sampling date	0	0
O5. Soil Surface Properties	Type of organic matter	0	0
O5. Soil Surface Properties	Dry mass of surface organic matter	0	0
O5. Soil Surface Properties	Nitrogen concentration in surface organic matter	0	0
O5. Soil Surface Properties	Phosphorus concentration in surface organic matter	0	0
O5. Soil Surface Properties	Potassium concentration in surface organic matter	0	0
O6. Soil Layer Properties	Experiment ID	0	0
O6. Soil Layer Properties	Site	0	0
O6. Soil Layer Properties	Year	0	0
O6. Soil Layer Properties	Treatment number	0	0
O6. Soil Layer Properties	Replicate	0	0
O6. Soil Layer Properties	Plot ID	0	0
O6. Soil Layer Properties	Sampling date	0	0
O6. Soil Layer Properties	Depth of measurement, top of soil layer	0	0

O6. Soil Layer Properties	Depth of measurement, bottom of soil layer	0	0
O6. Soil Layer Properties	Soil water content	0	0
O6. Soil Layer Properties	Nitrate N	0	0
O6. Soil Layer Properties	Ammonium N	0	0
O6. Soil Layer Properties	Total mineral N	0	0
O6. Soil Layer Properties	pH	0	0
O6. Soil Layer Properties	Cation exchange capacity	0	0
O6. Soil Layer Properties	Extractable P	0	0
O6. Soil Layer Properties	Potassium	0	0
O6. Soil Layer Properties	Magnesium	0	0
O6. Soil Layer Properties	Exchangeable Ca	0	0
O6. Soil Layer Properties	Potassium base saturation	0	0
O6. Soil Layer Properties	Magnesium base saturation	0	0
O6. Soil Layer Properties	Calcium base saturation	0	0
O6. Soil Layer Properties	Hydrogen base saturation	0	0
O7. Water	Experiment ID	0	0
O7. Water	Site	0	0
O7. Water	Year	0	0
O7. Water	Treatment number	0	0
O7. Water	Replicate	0	0
O7. Water	Plot ID	0	0
O7. Water	Sampling date	0	0
O7. Water	Sampling depth	0	0
O7. Water	Row position	0	0
O7. Water	Sampling method	0	0
O7. Water	NO3-N conc	0	0
O7. Water	Ammonium-N conc	0	0
O7. Water	Total Kjeldahl N conc	0	0
O7. Water	Water sample notes	0	0
S1. Soil Metadata	Soil ID	1	0
S1. Soil Metadata	Soil name	1	0
S1. Soil Metadata	Soil classification	1	0
S1. Soil Metadata	Soil classification system	1	0
S1. Soil Metadata	Source of soil data	1	0
S1. Soil Metadata	Latitude	1	0
S1. Soil Metadata	Longitude	1	0
S1. Soil Metadata	Elevation	1	0
S1. Soil Metadata	Anonymize	1	0
S1. Soil Metadata	Slope	0	1
S1. Soil Metadata	Soil surface color	0	1
S2. Soil Layer Properties	Soil ID	9	0
S2. Soil Layer Properties	Top of soil layer	9	0
S2. Soil Layer Properties	Bottom of soil layer	9	0
S2. Soil Layer Properties	Clay	9	0
S2. Soil Layer Properties	Silt	9	0
S2. Soil Layer Properties	Sand	9	0
S2. Soil Layer Properties	Gravel	9	0
S2. Soil Layer Properties	Organic matter	9	0
S2. Soil Layer Properties	Bulk density	9	0
S2. Soil Layer Properties	Wilting point	9	0
S2. Soil Layer Properties	Field capacity	9	0

S2. Soil Layer Properties	Saturated hydraulic conductivity	9	0
W1. Weather Station Metadata	Weather station ID	1	0
W1. Weather Station Metadata	Weather station name	1	0
W1. Weather Station Metadata	Latitude of station	1	0
W1. Weather Station Metadata	Longitude of station	1	0
W1. Weather Station Metadata	Elevation of weather station	1	0
W1. Weather Station Metadata	Anonymize	1	0
W1. Weather Station Metadata	Weather station temperature sensor height	1	0
W1. Weather Station Metadata	Weather station link	0	1
W2. Daily Weather Data	Weather station ID	365	0
W2. Daily Weather Data	Date	365	0
W2. Daily Weather Data	Minimum daily air temperature	365	0
W2. Daily Weather Data	Maximum daily air temperature	365	0
W2. Daily Weather Data	Daily precipitation	365	0
W2. Daily Weather Data	Solar radiation	365	0
W2. Daily Weather Data	Temperature, dewpoint	0	365
W2. Daily Weather Data	Wind speed, daily	0	365
W2. Daily Weather Data	PAR	365	0

3.0. Coherent Identifiers?

Index variables ('keys' in database terminology) from pairs of data frames are compared to make sure that the index values are identical across the sheets. This is fundamental to allowing different types of data to be linked across sheets. For example the values of 'Field location' should be the same in the sheets 'E1. Treatments' and 'E2. Fields'.

The basic approach for testing:

1. Create two temporary data frames.
2. Merge the data frames based on identifiers given as a list in the argument 'TestVar'.
3. Reduce the two data frames to just the columns corresponding to 'TestVar'.
4. Extract the unique combinations of values for each data frame.
5. Add flag variables, 'from_df1' and 'from_df2', to make it easier to detect problems.
6. Merge the the two data frames to create 'dfMerged'.
7. Compare the length of the two data frames. The lengths should be identical.
8. Print the merged test dataset 'dfMerged' to allow inspection by the users.

If the two frames are of different lengths, then there is a problem. If the two data frames are of the same length, one should still review 'from_df1' and 'from_df2' to see whether there are mismatches, which would be indicated by 'NA' in one of the two columns.

Common sources of mismatches include:

- Inconsistent use of spaces such as 'Blk 1' vs. 'Blk1' or other puncutation ('Blk_1').
- Simple spelling errors ('Fred' vs. 'Frred')
- Experiments, treatments or plots that were either never planted or not harvested.
- Extra rows being read in a given sheet, leading to an empty cell being assigned a value of NA. This may arise if a stray character appears outside of the intended range of data.

In the third case, it is helpful to provide a comment or note in the appropriate sheets to explain why additional data are missing.

3.1. Comparing identifiers used in M1. Experiments, E1. Treatments, E2. Fields and E3. Plots

[1] The sheets M1__Experiments and E1__Treatments have the same length

Experiment ID	From_df1	From_df2
UFGA8201	1	1

[1] The sheets E1__Treatments and E2__Fields have the same length

Experiment ID	From_df1	From_df2
UFGA8201	1	1

[1] The sheets E1__Treatments and E2__Fields have the same length

Experiment ID	Site	From_df1	From_df2
UFGA8201	UFGA	1	1

[1] The sheets E1__Treatments and E2__Fields have the same length

Experiment ID	Site	Field location	From_df1	From_df2
UFGA8201	UFGA	1	1	1

[1] The sheets E1__Treatments and E3__Plots have the same length

Experiment ID	Site	Field location	From_df1	From_df2
UFGA8201	UFGA	1	1	1

3.2. Comparing identifiers used in E2. Fields vs. E3. Plots

[1] The sheets E2__Fields and E3__Plots have the same length

Field location	From_df1	From_df2
1	1	1

[1] The sheets E2__Fields and E3__Plots have the same length

Experiment ID	Field location	From_df1	From_df2
UFGA8201	1	1	1

3.3. Comparing identifiers used for soil and weather data

Note that the same soil profile or weather data may be used for several experiments or nearby sites.

[1] The sheets E2__Fields and S1__Soil_Metadata have the same length

Soil ID	From__df1	From__df2
IBPN910015	1	1

[1] The sheets S1__Soil_Metadata and S2__Soil_Layer_Properties have the same length

Soil ID	From__df1	From__df2
IBPN910015	1	1

[1] The sheets E2__Fields and W1__Weather_Station_Metadata have the same length

Weather station ID	From__df1	From__df2
UFGA	1	1

[1] The sheets W1__Weather_Station_Metadata and W2__Daily_Weather_Data have the same length

Weather station ID	From__df1	From__df2
UFGA	1	1

3.4. Comparing identifiers in E1___Treatments and the various management sheets

Testing for matches is extended to sheets for irrigations, fertilizers, etc. Because not all sheets will have data, we first create a list of sheets with data (number of rows > 0).

Comparing E1__Treatments to E3__Plots for Experiment ID
[1] The sheets E1__Treatments and test_df have the same length

Comparing E1__Treatments to E4__Crop_Information for Experiment ID
[1] The sheets E1__Treatments and test_df have the same length

Comparing E1__Treatments to E5__Planting for Experiment ID
[1] The sheets E1__Treatments and test_df have the same length

Comparing E1__Treatments to E6__Irrigation for Experiment ID
[1] The sheets E1__Treatments and test_df have the same length

Comparing E1__Treatments to E7__Fertilizer for Experiment ID
[1] The sheets E1__Treatments and test_df have the same length

Comparing E1__Treatments to E9__Tillage for Experiment ID

```
[1] The sheets  E1__Treatments  and  test_df  have the same length

Comparing E1__Treatments to E10__Chemical_Applications for Experiment ID
[1] The sheets  E1__Treatments  and  test_df  have the same length

Comparing E1__Treatments to E11__Harvest for Experiment ID
[1] The sheets  E1__Treatments  and  test_df  have the same length
```

4. Compatible?: Checking That Variables Are Properly Described And Linked

We check whether all variables given in the various sheets appear in one of the three dictionary sheets. The dictionaries include variable names and definitions from the ICASA standards, so correct matching is needed to allow a dataset to be read by tools that use the ICASA standards.

A common source of mismatches is when a variable is added to crop or soil measurements but is not added in the dictionary sheets. When variables are present in both sources, possible causes of mismatches include:

- Differences in capitalization or punctuation
- Names with leading or trailing blank spaces

We also check whether all variables have definitions and are linked to ICASA short names. The processing works from the list of data frames, ls_sheets, but excludes the first three sheets and the three dictionaries.

4.1. Comparing the number of variables either used in the sheets or defined in the dictionaries

The initial check is whether the data sheets have roughly the same number of variables as the three dictionary sheets.

```
Total variables in the spreadsheet:  305
```

```
Total variables in the three dictionaries:  305
```

4.2. Compare lists of variables used in data sheets vs. the dictionaries

The second, more extensive check uses variable-by-variable matching. Mismatched variables are listed below. The columns InVUsed ("Included in Variables Used") and InDict ("In the Dictionaries") have a value of 1 if the variable is present in the respective source, the data sheets or the dictionaries. A value of NA means there is a mismatch.

The script displays only mismatches, 'VariableName' is truncated to 35 characters so that each comparison will appear on a single line.

SheetName	VariableName	InVUsed	InDict
O2. Yield Summary	Seed nitrogen conc	1	0
O2. Yield Summary	Seed nitrogen conc.	0	1

4.3. Checking whether all variables used in the data sheets are defined.

If present, the list below contains all variables that lack a definition. - If 'var_defined' = 0, the variable is in the Dictionary but lacks a definition. - If 'var_defined' = NA, the variable is not in the Dictionary_

SheetName	VariableName	var_defined
O2. Yield Summary	Seed nitrogen conc	NA

4.4. Checking whether all variables are linked to an ICASA short name.

If present, the list below contains all variables that are *not* associated with an ICASA short name. - If 'var_defined' = 0, the variable is in the Dictionary but no ICASA name is given. - If 'var_defined' = NA, the variable is not in the Dictionary_ Further information on ICASA variables may be found at the GitHub site.

SheetName	VariableName	has_ICASA_short
E7. Fertilizer	Analysis	0
E7. Fertilizer	Application timing	0
O1. Analysis Methods	Analysis method	0
O1. Analysis Methods	Analytical laboratory	0
O1. Analysis Methods	Computation method	0
O1. Analysis Methods	EPA method	0
O1. Analysis Methods	Matrix	0
O2. Yield Summary	Seed nitrogen conc	NA
O6. Soil Layer Properties	Calcium base saturation	0
O6. Soil Layer Properties	Hydrogen base saturation	0
O6. Soil Layer Properties	Magnesium base saturation	0
O6. Soil Layer Properties	Potassium base saturation	0
O7. Water	Row position	0
O7. Water	Sampling depth	0
O7. Water	Sampling method	0
O7. Water	Total Kjeldahl N conc	0
O7. Water	Water sample notes	0
S1. Soil Metadata	Anonymize	0
W1. Weather Station Metadata	Anonymize	0

4.5. Checking the Workbook for Formulas, Merged Cells or Commented Cells

One concern with use of spreadsheets is that , merged cells, comments attached to specific cells, or other features might cause problems in subsequent use of the data. We test first for use of formulas and merged cells, then test for comments attached to specific cells. The checking script only returns the cell address (e.g., 'B17') or range ('B5:C2'). To save space in the report, only first 20 cases are displayed.

4.5.1. Checking spreadsheet for formulae or merged cells

Use of formulas is dangerous in datasets that are redistributed because they may results in values being updated incorrectly.

When read by software expecting complete rows and columns, values of merged blocks of cells are typically assigned only to the upper left cell of a merged block, and other cells are assumed to have missing values. To avoid possible misinterpretation of data, all merged cells should be un-merged.

```
> For  E3. Plots  merged cells found at:
[1] "I1:K1,"
```

 (Only the first 20 cell ranges are displayed.)

4.5.2. Checking for cells with attached comment

If specific comments are attached to cells, the information may be lost in subsequent processing. The preferred way to record comments is in note or comment variables on the respective sheet.

If no sheets are listed above, then no comments attached to cells were found.

End of analysis for FDACS_UFGA8201_peanut.94.xlsx

Please post questions or feedback to https://github.com/DSSAT/Florida_Crop_BMP_Datasets/tree/main

Users who are familiar with R and Rmarkdown are encouraged to modify the script as needed.