FZJ Data Preparation

Elise

2024-06-09

Set the right working directory.

```
setwd("C:/Users/elise/Documents/Mémoire/Main/Data/Drive/FZJ")
```

Packages importation

1. Data importation

The first step in this data preparation process involves importing all the pertinent datasets listed in the Google Sheets "Variables template" document. Fist we find the files, then import them.

```
## [1] "env_multisensor.txt"    "FZJ.Rmd"              "FZJ_Initial Code Draft"
## [4] "img_extracted.txt"      "img_extracted_2.txt"  "img_extracted_3.txt"
## [7] "ISA_EPPN2020_FZJ.xlsx"  "multisensor.txt"
```

We can extract the coordinates of each plant with the ISA_EPPN.xlsx dataset, using a made-up function "coordinates_isaTAB".

```
# Get the coordinates
isaTAB <- read_excel("ISA_EPPN2020_FZJ.xlsx", sheet = "s_exp")
```

```
## New names:
## • `Unit` -> `Unit...9`
## • `Term Source REF` -> `Term Source REF...10`
## • `Term Accession Number` -> `Term Accession Number...11`
## • `Unit` -> `Unit...13`
## • `Term Source REF` -> `Term Source REF...14`
## • `Term Accession Number` -> `Term Accession Number...15`
## • `Unit` -> `Unit...22`
## • `Term Source REF` -> `Term Source REF...23`
## • `Term Accession Number` -> `Term Accession Number...24`
## • `Unit` -> `Unit...26`
## • `Term Source REF` -> `Term Source REF...27`
## • `Term Accession Number` -> `Term Accession Number...28`
```

```
coordinates <- coordinates_isaTAB(isaTAB)
```

A. Datasets structures

We can take a quick look at all the datasets.

- coordinates
- data
- data_environment

```
head(coordinates)
```

```
##   Sample.Name nrow ncol rep
## 1         101    1    1   5
## 2         102    2    1   5
## 3         103    3    1   5
## 4         104    4    1   5
## 5         105    5    1   5
## 6         106    6    1   5
```

```
head(data)
```

```
##   Unit.ID   Timestamp      Device.ID         Variable.ID Value
## 1     101 2019_11_11 MT_XS4002S_#1 Shoot_fresh_weight  6.61
## 2     102 2019_11_11 MT_XS4002S_#1 Shoot_fresh_weight  7.39
## 3     103 2019_11_11 MT_XS4002S_#1 Shoot_fresh_weight  4.60
## 4     104 2019_11_11 MT_XS4002S_#1 Shoot_fresh_weight  2.11
## 5     105 2019_11_11 MT_XS4002S_#1 Shoot_fresh_weight  7.56
## 6     106 2019_11_11 MT_XS4002S_#1 Shoot_fresh_weight  4.08
```

```
head(data_environment)
```

```
##                                 Unit.ID          Timestamp           Device.ID
## 1 EPPN2020_JRA1.4_OBJ2_GrowScreen-Rhizo1 23.10.2019 00:00 Comp.01_AHTR-R1 MY02
## 2 EPPN2020_JRA1.4_OBJ2_GrowScreen-Rhizo1 23.10.2019 00:30 Comp.01_AHTR-R1 MY02
## 3 EPPN2020_JRA1.4_OBJ2_GrowScreen-Rhizo1 23.10.2019 01:00 Comp.01_AHTR-R1 MY02
## 4 EPPN2020_JRA1.4_OBJ2_GrowScreen-Rhizo1 23.10.2019 01:30 Comp.01_AHTR-R1 MY02
## 5 EPPN2020_JRA1.4_OBJ2_GrowScreen-Rhizo1 23.10.2019 02:00 Comp.01_AHTR-R1 MY02
## 6 EPPN2020_JRA1.4_OBJ2_GrowScreen-Rhizo1 23.10.2019 02:30 Comp.01_AHTR-R1 MY02
##              Variable.ID    Value
## 1 Air Humidity (HMP110) 61.56047
## 2 Air Humidity (HMP110) 62.59469
## 3 Air Humidity (HMP110) 63.50576
## 4 Air Humidity (HMP110) 64.11512
## 5 Air Humidity (HMP110) 65.03021
## 6 Air Humidity (HMP110) 64.83519
```

B. Data manipulation

This next step standardizes diverse datasets by renaming variables for consistency, converting data into appropriate units, adding necessary columns, and merging the datasets.

```r
################################################################################
# COORDINATES
################################################################################
# Unit.ID
coordinates$Unit.ID <- seq_len(nrow(coordinates))
# Reference for Sample.Name et Unit.ID
reference <- coordinates[, c("Sample.Name", "Unit.ID")]
## We can then copy dataset2$Unit.ID <- reference$Unit.ID[match(dataset2$Sample.Name, r
eference$Sample.Name)]
# Genotype
reference$Genotype <- isaTAB$`Source Name`


################################################################################
# DATA
################################################################################
# separate for endpoint and timeseries
FW_shoot_g <- data[data$Variable.ID == "Shoot_fresh_weight", ]
FW_shoot_g$Date <- as.Date("2019-11-11")
FW_shoot_g$unit <- reference$Unit.ID[match(FW_shoot_g$Unit.ID, reference$Sample.Name)]
FW_shoot_g$Unit.ID <- FW_shoot_g$unit

DW_shoot_g <- data[data$Variable.ID == "Shoot_dry_weight", ]
DW_shoot_g$Date <- as.Date("2019-11-11")
DW_shoot_g$unit <- reference$Unit.ID[match(DW_shoot_g$Unit.ID, reference$Sample.Name)]
DW_shoot_g$Unit.ID <- DW_shoot_g$unit

DW_root_g <- data[data$Variable.ID == "Root_dry_weight", ]
DW_root_g$Date <- as.Date("2019-11-11")
DW_root_g$unit <- reference$Unit.ID[match(DW_root_g$Unit.ID, reference$Sample.Name)]
DW_root_g$Unit.ID <- DW_root_g$unit

Root_length_cm <- data[data$Variable.ID == "Total_root_length_scan", ]
Root_length_cm$Date <- as.Date("2019-11-11")
Root_length_cm$unit <- reference$Unit.ID[match(Root_length_cm$Unit.ID, reference$Sampl
e.Name)]
Root_length_cm$Unit.ID <- Root_length_cm$unit

Root_angle <- data[data$Variable.ID == "Root_angle", ]
Root_angle$Date <- as.Date("2019-11-11")
Root_angle$unit <- reference$Unit.ID[match(Root_angle$Unit.ID, reference$Sample.Name)]
Root_angle$Unit.ID <- Root_angle$unit

Root_number <- data[data$Variable.ID == "Root_number", ]
Root_number$Date <- as.Date("2019-11-11")
Root_number$unit <- reference$Unit.ID[match(Root_number$Unit.ID, reference$Sample.Nam
e)]
Root_number$Unit.ID <- Root_number$unit


Leaf_number <- data[data$Variable.ID == "Leaf_number", ]
Leaf_number$Date <- gsub("_", "-", Leaf_number$Timestamp)
Leaf_number$Date <- as.Date(Leaf_number$Date)
Leaf_number$unit <- reference$Unit.ID[match(Leaf_number$Unit.ID, reference$Sample.Nam
e)]
Leaf_number$Unit.ID <- Leaf_number$unit
```

```
Plant_height_cm <- data[data$Variable.ID == "Plant_height", ]
Plant_height_cm$Date <- gsub("_", "-", Plant_height_cm$Timestamp)
Plant_height_cm$Date <- as.Date(Plant_height_cm$Date)
Plant_height_cm$unit <- reference$Unit.ID[match(Plant_height_cm$Unit.ID, reference$Samp
le.Name)]
Plant_height_cm$Unit.ID <- Plant_height_cm$unit
```

2. Data template

A. Data template: plant_info

This dataset contains information about the plant: Unit.ID, genotype, replication, row and column location in the greenhouse, and soil treatment.

B. Data template: endpoint

This datasets contains information of the end of the experiment (variables at harvest). It is then linked by the Unit.ID to the plant_info data template.

C. Data template: timeseries

This section in divided in three data templates:

- timeseries
- S_timeseries (variables computed from sideview imaging or image processing)
- T_timeseries (variables computed from topview imaging or image processing)

The time interval between data timestamps varies in each platform. They are then linked by the Unit.ID to the plant_info data template.

D. FZJ data templates

- plant_info
- endpoint
- timeseries
- S_timeseries
- T_timeseries

```
##    Unit.ID Genotype Soil Replication Row Column Platform
## 1        1  EPPN2_H   NA           5   1      1      FZJ
## 2        2  EPPN4_H   NA           5   2      1      FZJ
## 3        3  EPPN2_L   NA           5   3      1      FZJ
## 4        4 EPPN20_T   NA           5   4      1      FZJ
## 5        5  EPPN1_L   NA           5   5      1      FZJ
## 6        6  EPPN4_L   NA           5   6      1      FZJ
```

```
##   Unit.ID Time       Date Timestamp DW_shoot_g FW_shoot_g DW_root_g FW_root_g
## 1       1   NA 2019-11-11        NA     0.5506       6.61    0.0838        NA
## 2       2   NA 2019-11-11        NA     0.5655       7.39    0.0700        NA
## 3       3   NA 2019-11-11        NA     0.4032       4.60    0.0877        NA
## 4       4   NA 2019-11-11        NA     0.1335       2.11    0.0176        NA
## 5       5   NA 2019-11-11        NA         NA       7.56    0.1071        NA
## 6       6   NA 2019-11-11        NA     0.3114       4.08    0.0649        NA
##   Leaf_number Plant_height_cm DW_plant_g Root_length_cm Root_number Root_angle
## 1          NA              NA         NA        2137.90          13         69
## 2          NA              NA         NA        1622.61          11         67
## 3          NA              NA         NA        1766.15           6        128
## 4          NA              NA         NA         275.45           4         11
## 5          NA              NA         NA        2122.43           8         82
## 6          NA              NA         NA        1170.97           9        109
##   Total_wu DW_seed_g FW_seed_g Leaf_area_cmsquared Genotype Soil Replication
## 1       NA        NA        NA                  NA   EPPN2_H   NA           5
## 2       NA        NA        NA                  NA   EPPN4_H   NA           5
## 3       NA        NA        NA                  NA   EPPN2_L   NA           5
## 4       NA        NA        NA                  NA  EPPN20_T   NA           5
## 5       NA        NA        NA                  NA   EPPN1_L   NA           5
## 6       NA        NA        NA                  NA   EPPN4_L   NA           5
##   Row Column Platform
## 1   1      1      FZJ
## 2   2      1      FZJ
## 3   3      1      FZJ
## 4   4      1      FZJ
## 5   5      1      FZJ
## 6   6      1      FZJ
```

```
##   Unit.ID Time       Date Timestamp Manual_Plant_height_cm Leaf_number Wue
## 1       1   NA 2019-10-28        NA                   10.0         2.2  NA
## 2       1   NA 2019-10-31        NA                   21.1         3.2  NA
## 3       1   NA 2019-11-04        NA                   33.4         4.0  NA
## 4       1   NA 2019-11-07        NA                   38.2         4.8  NA
## 5       1   NA 2019-11-11        NA                   45.5         5.8  NA
## 6       2   NA 2019-10-28        NA                    8.5         2.2  NA
##   Plant_biomass Ligulated_leaf_number Plant_emergence Plant_transpiration
## 1            NA                    NA              NA                  NA
## 2            NA                    NA              NA                  NA
## 3            NA                    NA              NA                  NA
## 4            NA                    NA              NA                  NA
## 5            NA                    NA              NA                  NA
## 6            NA                    NA              NA                  NA
##   Daily_wu Soil_water_potential Genotype Soil Replication Row Column Platform
## 1       NA                   NA  EPPN2_H   NA           5   1      1      FZJ
## 2       NA                   NA  EPPN2_H   NA           5   1      1      FZJ
## 3       NA                   NA  EPPN2_H   NA           5   1      1      FZJ
## 4       NA                   NA  EPPN2_H   NA           5   1      1      FZJ
## 5       NA                   NA  EPPN2_H   NA           5   1      1      FZJ
## 6       NA                   NA  EPPN4_H   NA           5   2      1      FZJ
```

```
##   Unit.ID Timestamp Date Time S_Height_cm S_Height_pixel S_Area_cmsquared
## 1    <NA>        NA   NA   NA          NA             NA               NA
##   S_Area_pixel S_Perimeter_cm S_Perimeter_pixel S_Convex_hull_area_cmsquared
## 1           NA             NA                NA                           NA
##   S_Solidity S_Compactness S_Width_cm S_Width_pixel S_Leaf_area_cmsquared
## 1         NA            NA         NA            NA                    NA
##   Genotype Soil Replication  Row Column Platform
## 1     <NA>   NA        <NA> <NA>   <NA>     <NA>
```

```
##   Unit.ID Time Date Timestamp T_Area_cm_squared T_Area_pixel T_Perimeter_cm
## 1    <NA>   NA   NA        NA                NA           NA             NA
##   T_Perimeter_pixel T_Convex_hull_area_cmsquared T_Solidity T_Compactness
## 1                NA                           NA         NA            NA
##   T_Roundness T_Roundness2 T_Isotropy T_Eccentricity T_Rms T_Sol Genotype Soil
## 1          NA           NA         NA             NA    NA    NA     <NA>   NA
##   Replication  Row Column Platform
## 1        <NA> <NA>   <NA>     <NA>
```

3. Export the data templates in .txt

Stock the new data sets in a new folder.

```
setwd("C:/Users/elise/Documents/Mémoire/Main/Data/Templates/FZJ")

write.table(plant_info, file = "plant_info.txt", sep = "\t", row.names = FALSE, quote =
FALSE)
write.table(endpoint, file = "endpoint.txt", sep = "\t", row.names = FALSE, quote = FAL
SE)
write.table(timeseries, file = "timeseries.txt", sep = "\t", row.names = FALSE, quote =
FALSE)
write.table(S_timeseries, file = "S_timeseries.txt", sep = "\t", row.names = FALSE, quo
te = FALSE)
write.table(T_timeseries, file = "T_timeseries.txt", sep = "\t", row.names = FALSE, quo
te = FALSE)
```