UCPH Data Preparation

Elise

2024-06-09

Set the right working directory.

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

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] "2021-06 TNA - Mspec v1.0 0001.csv"
##  [2] "Climate data_2021.06.02._2021_06.13_CrossPlatform Cell 3.xls"
##  [3] "Climate data_2021.06.13._2021.07.01_CrossPlatform Phenolab,.xls"
##  [4] "Cross Platform_Automatic_Height.xlsx"
##  [5] "DesignDIGGER_PhenoLab_JointPf.csv"
##  [6] "EPPN_Biomass_PhenoLab.xlsx"
##  [7] "Height measurements last 2 days.xlsm"
##  [8] "ISA_EPPN2020_UCPH.xlsx"
##  [9] "UCPH_Data-Preparation.html"
## [10] "UCPH_Data Preparation.Rmd"
## [11] "UCPH_Initial Code Draft"
## [12] "UCPH_Template"
```

We can extract the coordinates of each plant with the ISA_EPPN.xlsx dataset, using a made-up function “coordinates_isaTAB”. For UCPH, the coordinates are not in the ISA_EPPN dataset.

```
design <- read.csv("DesignDIGGER_PhenoLab_JointPf.csv")
```

A. Datasets structures

We can take a quick look at all the datasets.

- design
- data_pheno
- data_imaging

```
head(design)
```

```
##   UNIT ROW RANGE REP Substrate Genotype
## 1    1   1     1   1        S2  EPPN2_H
## 2    2   2     1   1        S1   EPPN_T
## 3    3   3     1   2        S2  EPPN1_H
## 4    4   4     1   2        S1  EPPN2_H
## 5    5   5     1   3        S2  EPPN1_L
## 6    6   6     1   3        S1  EPPN3_L
```

```
head(data_pheno)
```

```
## # A tibble: 6 × 4
##   Fixture_number Maize_cultivar `DW_shoot(g)`      `DW_root(g)`
##            <dbl> <chr>          <chr>                     <dbl>
## 1              1 S2,EPPN1_L     2.1916279069767439          1
## 2              2 S2,EPPN4_H     2.7155378486055777          0.5
## 3              3 S2,EPPN1_H     2.7084942084942085          0.3
## 4              4 S2,EPPN2_L     1.1372549019607843          0.1
## 5              5 S2,EPPN3_L     2.1572052401746724          0.4
## 6              6 S2,EPPN4_L     1.3064516129032258          0.4
```

```
head(data_imaging)
```

```
## # A tibble: 6 × 4
##   `Date and Time`     `Type:`     `Fixture ID:` `Measurement:`
##   <dttm>              <chr>               <dbl>          <dbl>
## 1 2021-06-29 05:25:00 Height (mm)           102            583
## 2 2021-06-29 05:23:17 Height (mm)            93            379
## 3 2021-06-29 05:21:38 Height (mm)            84            516
## 4 2021-06-29 05:20:09 Height (mm)            75            606
## 5 2021-06-29 05:18:44 Height (mm)            66            606
## 6 2021-06-29 05:17:26 Height (mm)            57            564
```

```
#head(data_environment)
```

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.

```
################################################################################
# DESIGN
################################################################################
# Unit.ID
design <- rename(design,
                 Unit.ID = UNIT,
                 nrow = ROW,
                 ncol = RANGE,
                 rep = REP,
                 Soil = Substrate)

################################################################################
# DATA_PHENO
################################################################################
# Time, Date and Timestamp
data_pheno$Date <- as.Date("2021-06-28")

# Name of the platform
data_pheno$Platform <- "UCPH"

# Rename the columns for the template
data_pheno <- rename(data_pheno,
                     Unit.ID = Fixture_number,
                     DW_shoot_g = `DW_shoot(g)`,
                     DW_root_g = `DW_root(g)`
                     )

################################################################################
# DATA_IMAGING
################################################################################
# Time, Date and Timestamp
data_imaging$Timestamp <- data_imaging$`Date and Time`

data_imaging$Date <- sapply(strsplit(as.character(data_imaging$Timestamp), split = "
"), '[', 1)
data_imaging$Time <- sapply(strsplit(as.character(data_imaging$Timestamp), split = "
"), '[', 2)

data_imaging <- subset(data_imaging, Date >= "2021-06-13") # The plants were put in the
platform on the 13rd of june



# Name of the platform
data_imaging$Platform <- "UCPH"

# Rename the columns for the template
data_imaging <- rename(data_imaging,
                       Unit.ID = `Fixture ID:`,
                       Plant_height_mm = `Measurement:`
                           )
```

Unit conversions

The data template is only in cm, cm² and g. This step converts the data in the right units.

For the UCPH platform, one variable is in mm.

```
data_imaging$S_Height_cm <- 0.01 * data_imaging$Plant_height_mm
```

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. NaPPI data templates

- plant_info
- endpoint
- timeseries
- S_timeseries
- T_timeseries

```
##   Unit.ID Genotype Soil Replication Row Column Platform
## 1       1  EPPN2_H   S2           1   1      1     UCPH
## 2       2   EPPN_T   S1           1   2      1     UCPH
## 3       3  EPPN1_H   S2           2   3      1     UCPH
## 4       4  EPPN2_H   S1           2   4      1     UCPH
## 5       5  EPPN1_L   S2           3   5      1     UCPH
## 6       6  EPPN3_L   S1           3   6      1     UCPH
```

```
##   Unit.ID Time       Date Timestamp DW_shoot_g FW_shoot_g DW_root_g FW_root_g
## 1       1   NA 2021-06-28        NA   2.191628         NA       1.0        NA
## 2       2   NA 2021-06-28        NA   2.715538         NA       0.5        NA
## 3       3   NA 2021-06-28        NA   2.708494         NA       0.3        NA
## 4       4   NA 2021-06-28        NA   1.137255         NA       0.1        NA
## 5       5   NA 2021-06-28        NA   2.157205         NA       0.4        NA
## 6       6   NA 2021-06-28        NA   1.306452         NA       0.4        NA
##   Leaf_number Plant_height_cm DW_plant_g Root_length_cm Root_number Root_angle
## 1          NA              NA         NA             NA          NA         NA
## 2          NA              NA         NA             NA          NA         NA
## 3          NA              NA         NA             NA          NA         NA
## 4          NA              NA         NA             NA          NA         NA
## 5          NA              NA         NA             NA          NA         NA
## 6          NA              NA         NA             NA          NA         NA
##   Total_wu DW_seed_g FW_seed_g Leaf_area_cmsquared Genotype Soil Replication
## 1       NA        NA        NA                  NA  EPPN2_H   S2           1
## 2       NA        NA        NA                  NA   EPPN_T   S1           1
## 3       NA        NA        NA                  NA  EPPN1_H   S2           2
## 4       NA        NA        NA                  NA  EPPN2_H   S1           2
## 5       NA        NA        NA                  NA  EPPN1_L   S2           3
## 6       NA        NA        NA                  NA  EPPN3_L   S1           3
##   Row Column Platform
## 1   1      1     UCPH
## 2   2      1     UCPH
## 3   3      1     UCPH
## 4   4      1     UCPH
## 5   5      1     UCPH
## 6   6      1     UCPH
```

```
##   Unit.ID Time Date Timestamp Manual_Plant_height_cm Leaf_number Wue
## 1    <NA>   NA   NA        NA                     NA          NA  NA
##   Plant_biomass Ligulated_leaf_number Plant_emergence Plant_transpiration
## 1            NA                    NA              NA                  NA
##   Daily_wu Soil_water_potential Genotype Soil Replication  Row Column Platform
## 1       NA                   NA     <NA> <NA>        <NA> <NA>   <NA>     <NA>
```

```
##   Unit.ID           Timestamp       Date        Time S_Height_cm S_Height_pixel
## 1     102 2021-06-29 05:25:00 2021-06-29 05:25:00.55        5.83             NA
## 2      93 2021-06-29 05:23:17 2021-06-29 05:23:17.15        3.79             NA
## 3      84 2021-06-29 05:21:38 2021-06-29 05:21:38.45        5.16             NA
## 4      75 2021-06-29 05:20:09 2021-06-29 05:20:09.05        6.06             NA
## 5      66 2021-06-29 05:18:44 2021-06-29 05:18:44.45        6.06             NA
## 6      57 2021-06-29 05:17:26 2021-06-29 05:17:26.75        5.64             NA
##   S_Area_cmsquared S_Area_pixel S_Perimeter_cm S_Perimeter_pixel
## 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
##   S_Convex_hull_area_cmsquared S_Solidity S_Compactness S_Width_cm
## 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
##   S_Width_pixel S_Leaf_area_cmsquared Genotype Soil Replication Row Column
## 1            NA                    NA  EPPN2_H   S1           3   6      9
## 2            NA                    NA  EPPN2_H   S1           5   9      8
## 3            NA                    NA  EPPN3_H   S1           6  12      7
## 4            NA                    NA   EPPN_T   S2           2   3      7
## 5            NA                    NA  EPPN3_H   S1           3   6      6
## 6            NA                    NA  EPPN1_H   S1           5   9      5
##   Platform
## 1     UCPH
## 2     UCPH
## 3     UCPH
## 4     UCPH
## 5     UCPH
## 6     UCPH
```

```
##   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/UCPH")

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)
```