4PMI Data Analysis Timepoints

Elise

2024-06-09

Set the right working directory.

```
setwd("C:/Users/elise/Documents/Mémoire/Main/Data/Templates/4PMI")
```

Data importation

Reimport the data sets extracted from the Data Preparation and Data Analysis R Markdown.

```
list.files()
```

```
## [1] "endpoint.txt"   "plant_info.txt"
```

```
plant_info <- read.table("plant_info.txt", header = TRUE, sep = "\t")
endpoint <- read.table("endpoint.txt", header = TRUE, sep = "\t")

# plant_info
plant_info <- lapply(plant_info, factor)

# endpoint
matching_cols <- intersect(names(endpoint), names(plant_info))
endpoint[, matching_cols] <- lapply(endpoint[, matching_cols], factor)
endpoint$Date <- date(endpoint$Date)
endpoint$Timestamp <- NA

platform <- "4PMI"

# endpoint
df <- endpoint[,colSums(is.na(endpoint))<nrow(endpoint)]
genotype_index <- which(colnames(df) == "Genotype")
variables <- colnames(df[, c(3:(genotype_index - 1))]) # We remove the two first column
s that are "Unit.ID" and "Date"

print(paste(platform, ": The variables for endpoint are", paste(variables, collapse =
", "), sep = " "))
```

```
## [1] "4PMI : The variables for endpoint are DW_shoot_g, DW_root_g, Leaf_number, DW_se
ed_g"
```

```
endpoint$Plant_type <- substr(endpoint$Genotype, nchar(as.character(endpoint$Genotyp
e)), nchar(as.character(endpoint$Genotype)))
```

Get the cleaned endpoint data

```
endpoint_clean <- endpoint
# Run the function on the dataset for all the variables
endpoint_clean <- detect_replace_outliers_by_genotype(endpoint_clean)
```

Time point objects

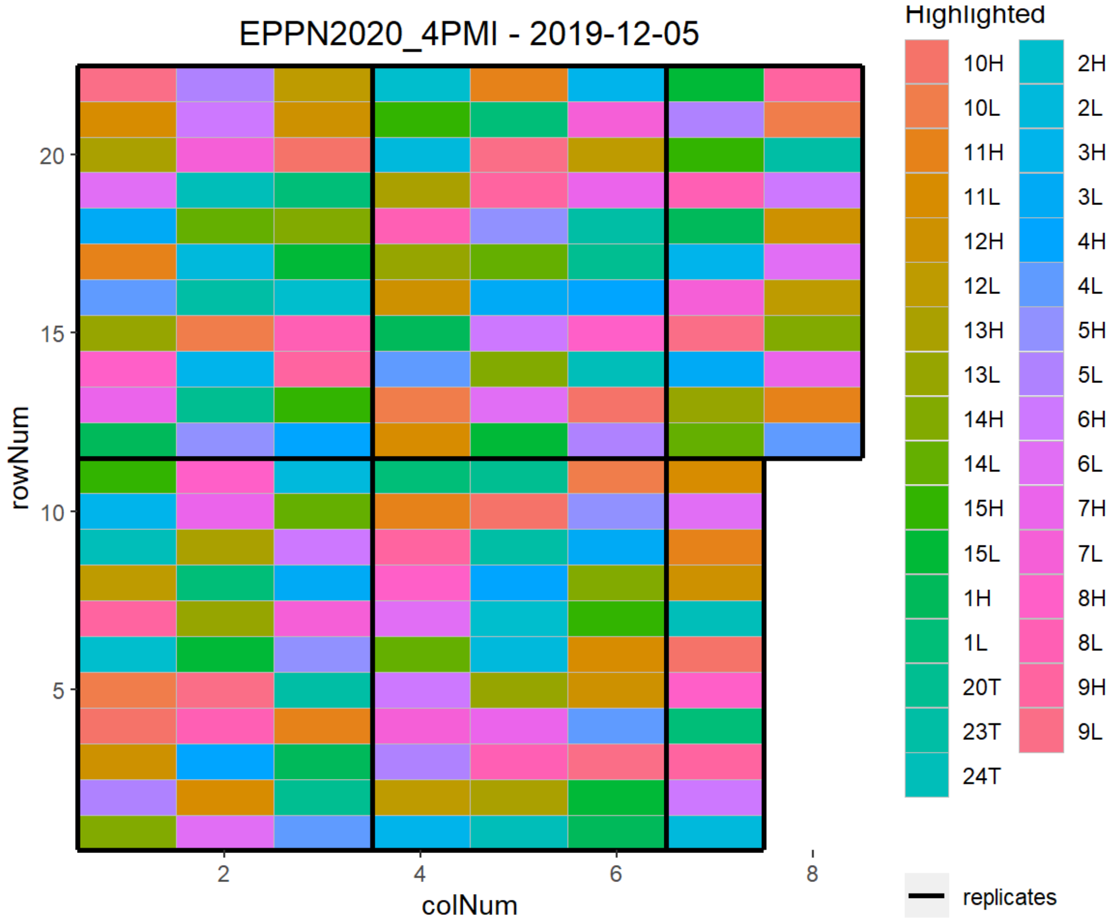
Generation of the timePoints objects using the function “createTimePoints”.

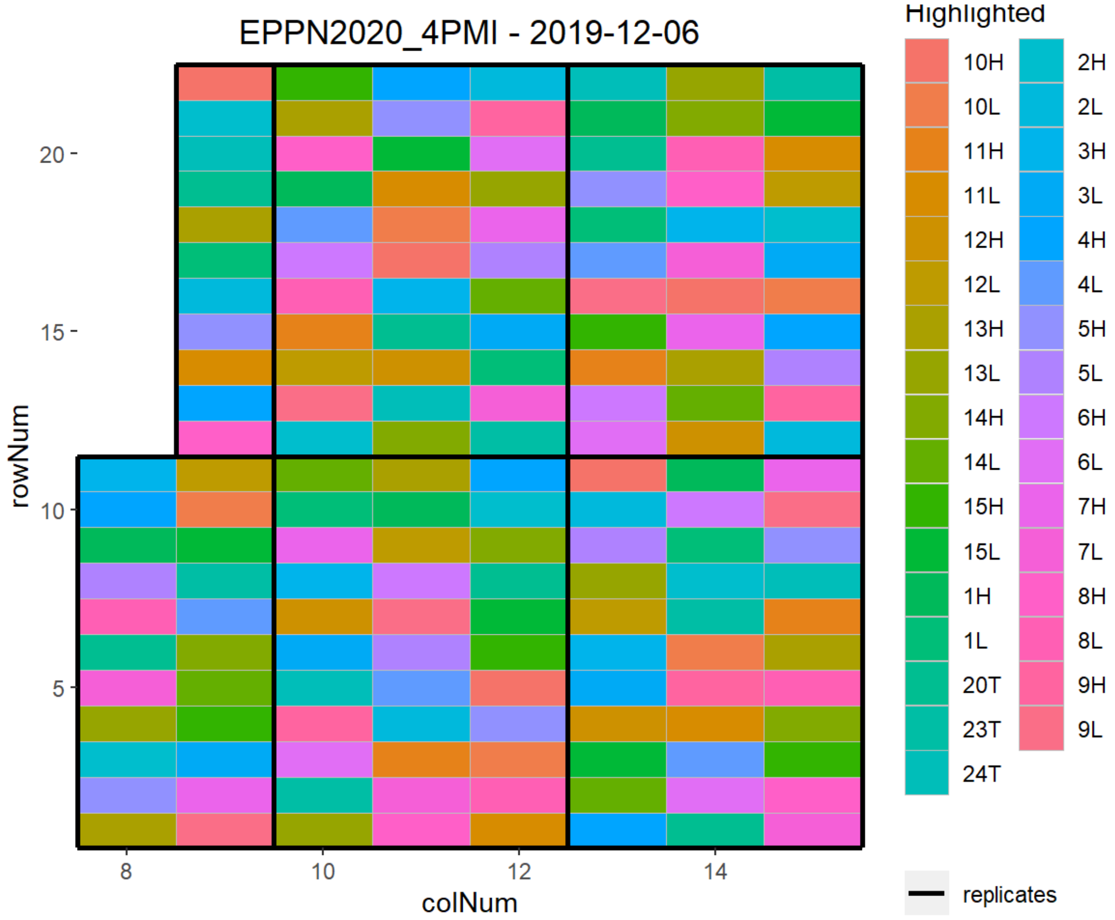
```
timePoint_endpoint <- createTimePoints(dat = endpoint,
                                      experimentName = "EPPN2020_4PMI",
                                      genotype = "Genotype",
                                      timePoint = "Date",
                                      plotId = "Unit.ID",
                                      rowNum = "Row",
                                      colNum = "Column",
                                      repId = "Replication")

timePoint_endpoint_clean <- createTimePoints(dat = endpoint_clean,
                                      experimentName = "EPPN2020_4PMI",
                                      genotype = "Genotype",
                                      timePoint = "Date",
                                      plotId = "Unit.ID",
                                      rowNum = "Row",
                                      colNum = "Column",
                                      repId = "Replication")
```

Gentoypic layout

Check the layout of the platforms’ genotypes.

```
genotypes_list <- as.character(unique(endpoint$Genotype))

plot(timePoint_endpoint,
     plotType = "layout",
     highlight =  genotypes_list,
     showGeno = FALSE)
```


EPPN2020_4PMI - 2019-12-05
rowNum
colNum
20
15
10
5
2
4
6
8
Highlighted
10H
10L
11H
11L
12H
12L
13H
13L
14H
14L
15H
15L
1H
1L
20T
23T
24T
2H
2L
3H
3L
4H
4L
5H
5L
6H
6L
7H
7L
8H
8L
9H
9L
replicates


EPPN2020_4PMI - 2019-12-06
rowNum
colNum
20
15
10
5
8
10
12
14
Highlighted
10H
10L
11H
11L
12H
12L
13H
13L
14H
14L
15H
15L
1H
1L
20T
23T
24T
2H
2L
3H
3L
4H
4L
5H
5L
6H
6L
7H
7L
8H
8L
9H
9L
replicates

1. endpoint

Comparisons between raw and cleaned data

View timePoint object.

```
summary(timePoint_endpoint)
```

```
## timePoint_endpoint contains data for experiment EPPN2020_4PMI.
## 
## It contains 2 time points.
## First time point: 2019-12-05
## Last time point: 2019-12-06
## 
## No check genotypes are defined.
```

```
getTimePoints(timePoint_endpoint)
```

```
##   timeNumber  timePoint
## 1          1 2019-12-05
## 2          2 2019-12-06
```

Count the number of observations per trait.

```
traits <- variables

for (trait_name in traits) {
  print(paste("How many data observations for", trait_name))
  num_observations <- countValid(timePoint_endpoint, trait_name)
  print(num_observations)
}
```

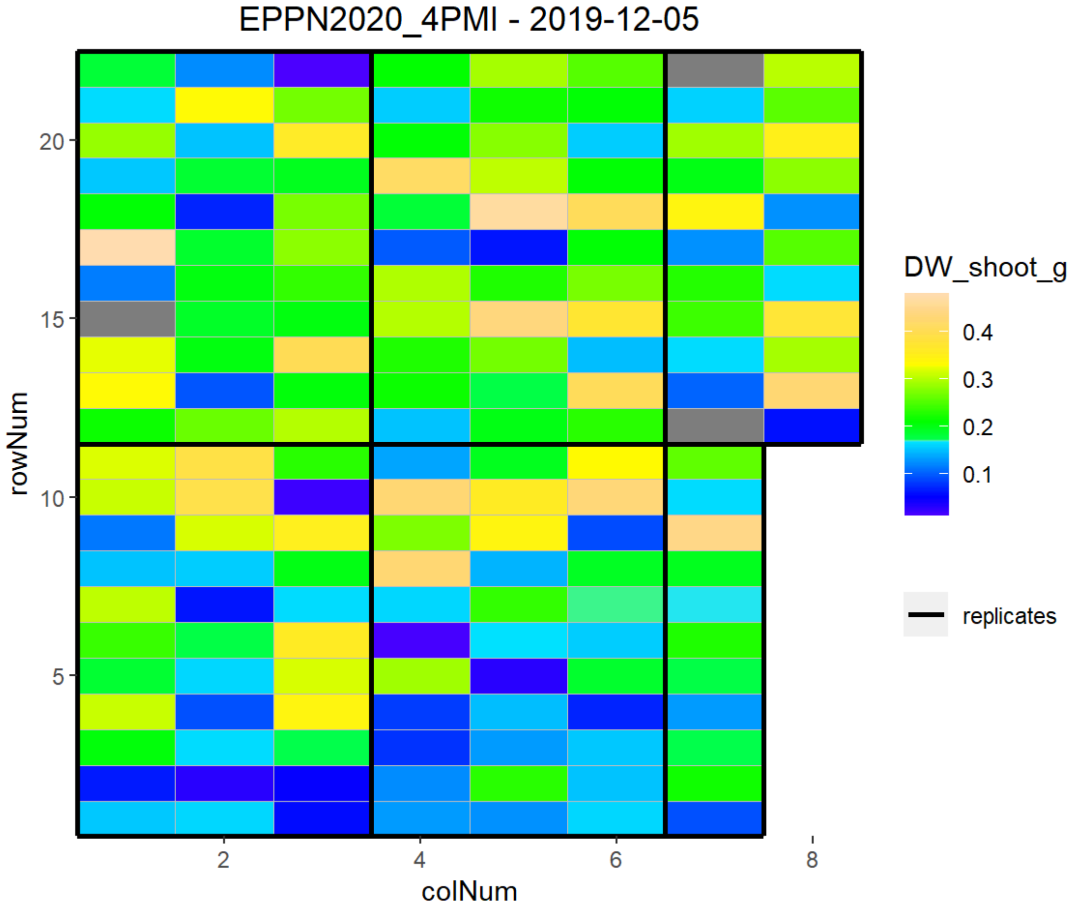
```
## [1] "How many data observations for DW_shoot_g"
## 2019-12-05 2019-12-06 
##        162        163 
## [1] "How many data observations for DW_root_g"
## 2019-12-05 2019-12-06 
##        162        162 
## [1] "How many data observations for Leaf_number"
## 2019-12-05 2019-12-06 
##        161        158 
## [1] "How many data observations for DW_seed_g"
## 2019-12-05 2019-12-06 
##        165        165
```

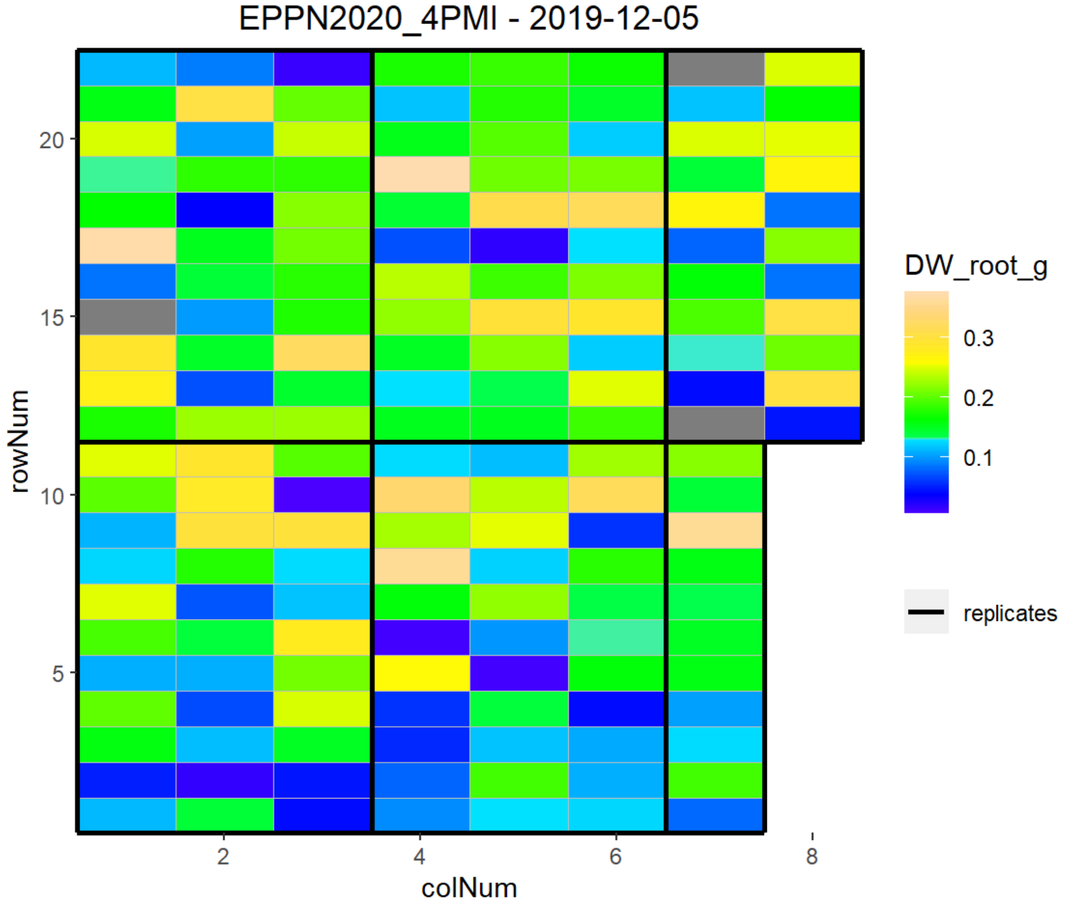
```
for (trait_name in traits) {
  print(paste("How many cleaned data observations for", trait_name))
  num_observations <- countValid(timePoint_endpoint_clean, trait_name)
  print(num_observations)
}
```

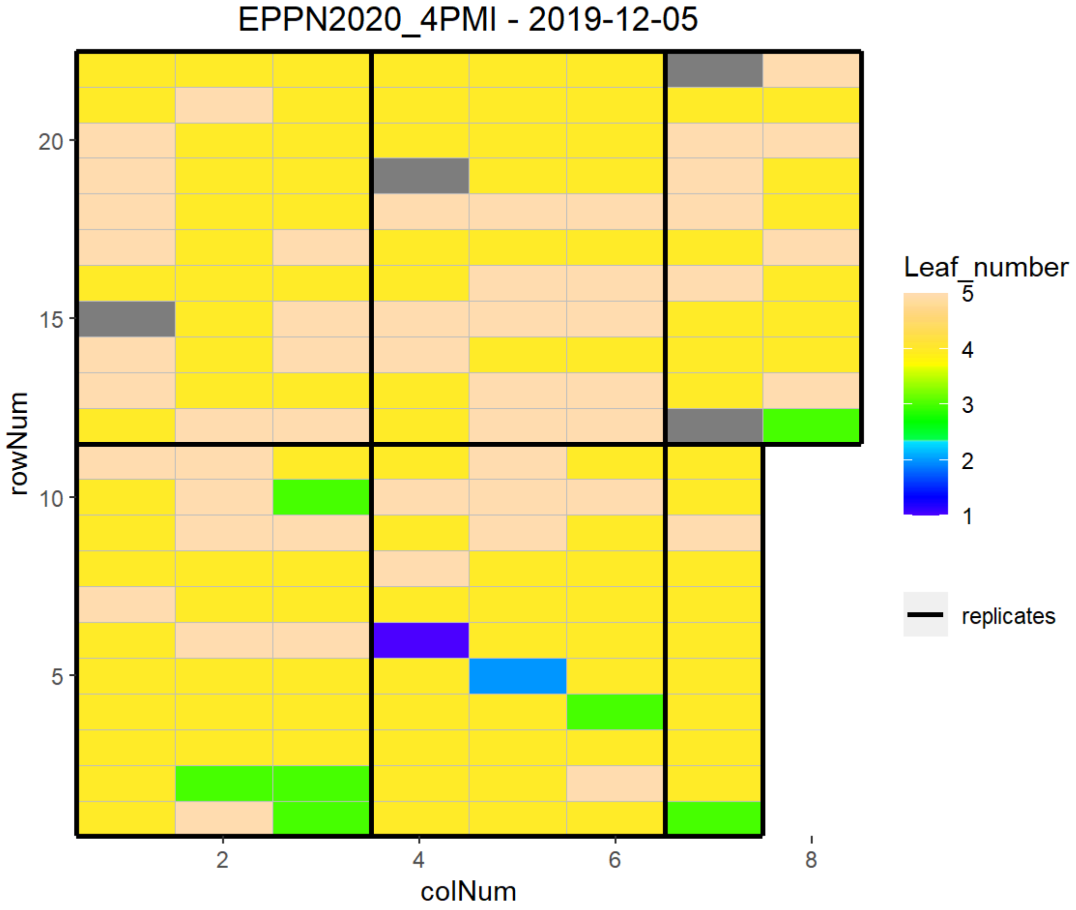
```
## [1] "How many cleaned data observations for DW_shoot_g"
## 2019-12-05 2019-12-06 
##        160        151 
## [1] "How many cleaned data observations for DW_root_g"
## 2019-12-05 2019-12-06 
##        157        157 
## [1] "How many cleaned data observations for Leaf_number"
## 2019-12-05 2019-12-06 
##        155        145 
## [1] "How many cleaned data observations for DW_seed_g"
## 2019-12-05 2019-12-06 
##        152        155 
```

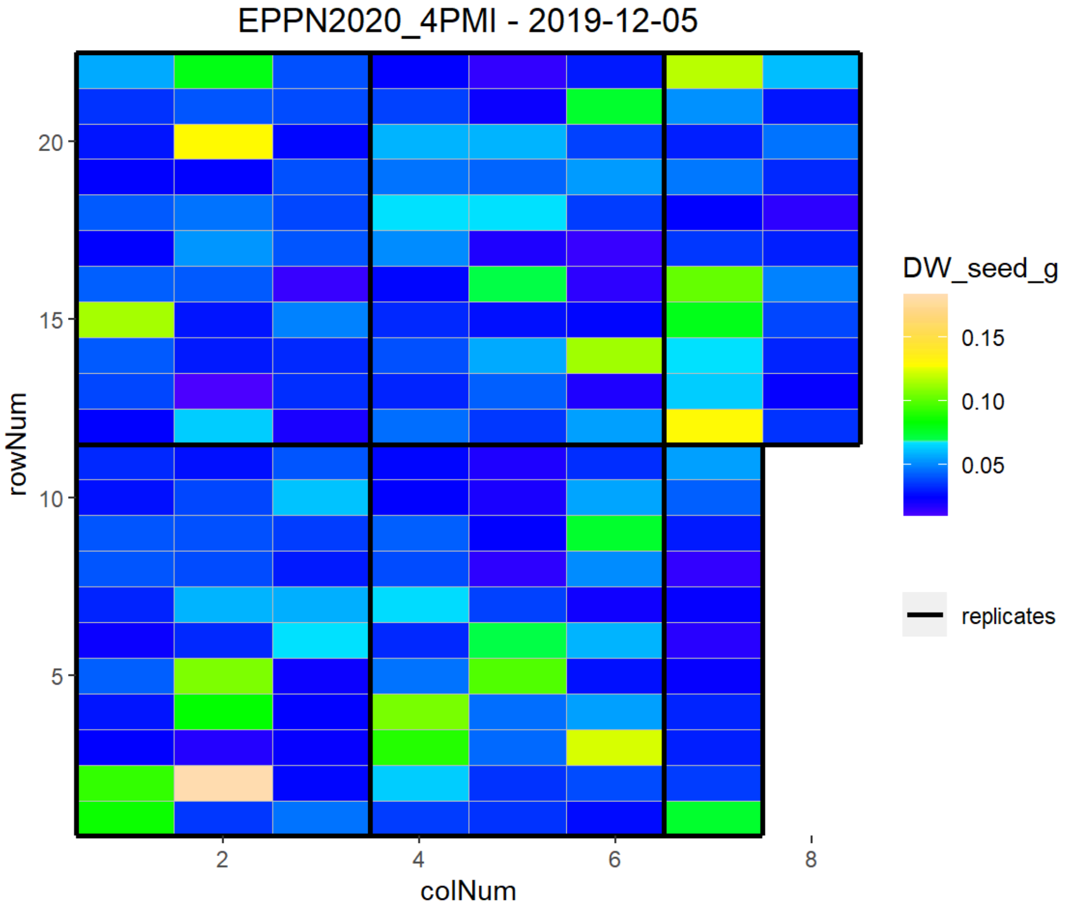
Check the heatmap of the data at harvest

```
for (trait_name in traits) {
  plot(timePoint_endpoint,
     plotType = "layout",
     timePoints = 1,
     traits = trait_name)
}
```

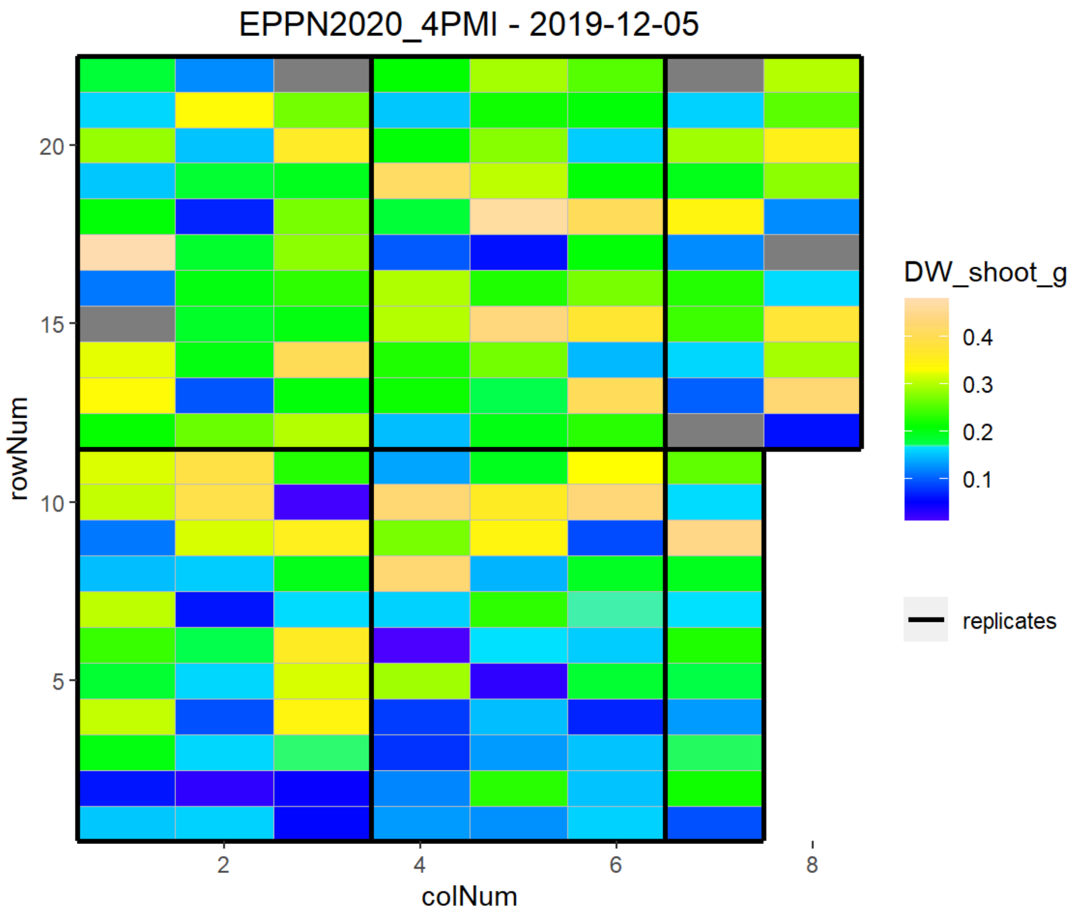

EPPN2020_4PMI - 2019-12-05
rowNum
20
15
10
5
colNum
2
4
6
8
DW_shoot_g
0.4
0.3
0.2
0.1
replicates

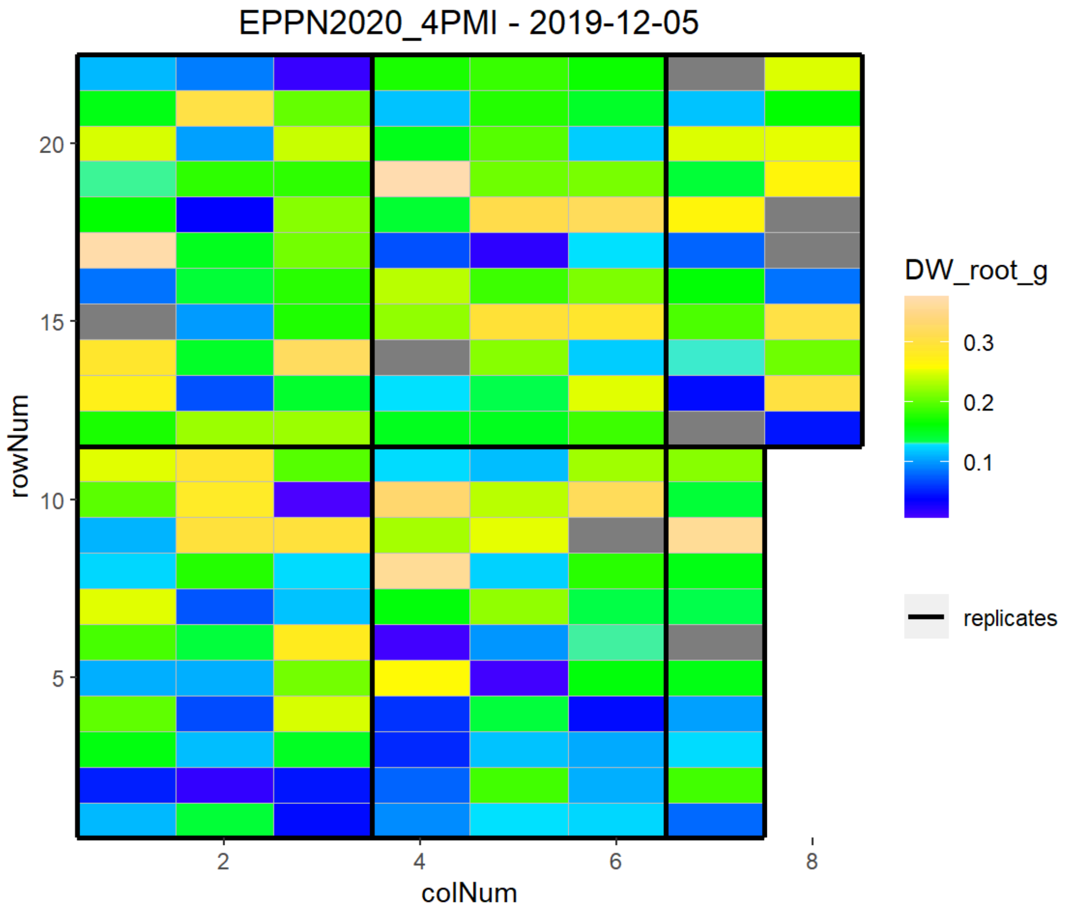

EPPN2020_4PMI - 2019-12-05
rowNum
20
15
10
5
colNum
2
4
6
8
DW_root_g
0.3
0.2
0.1
replicates

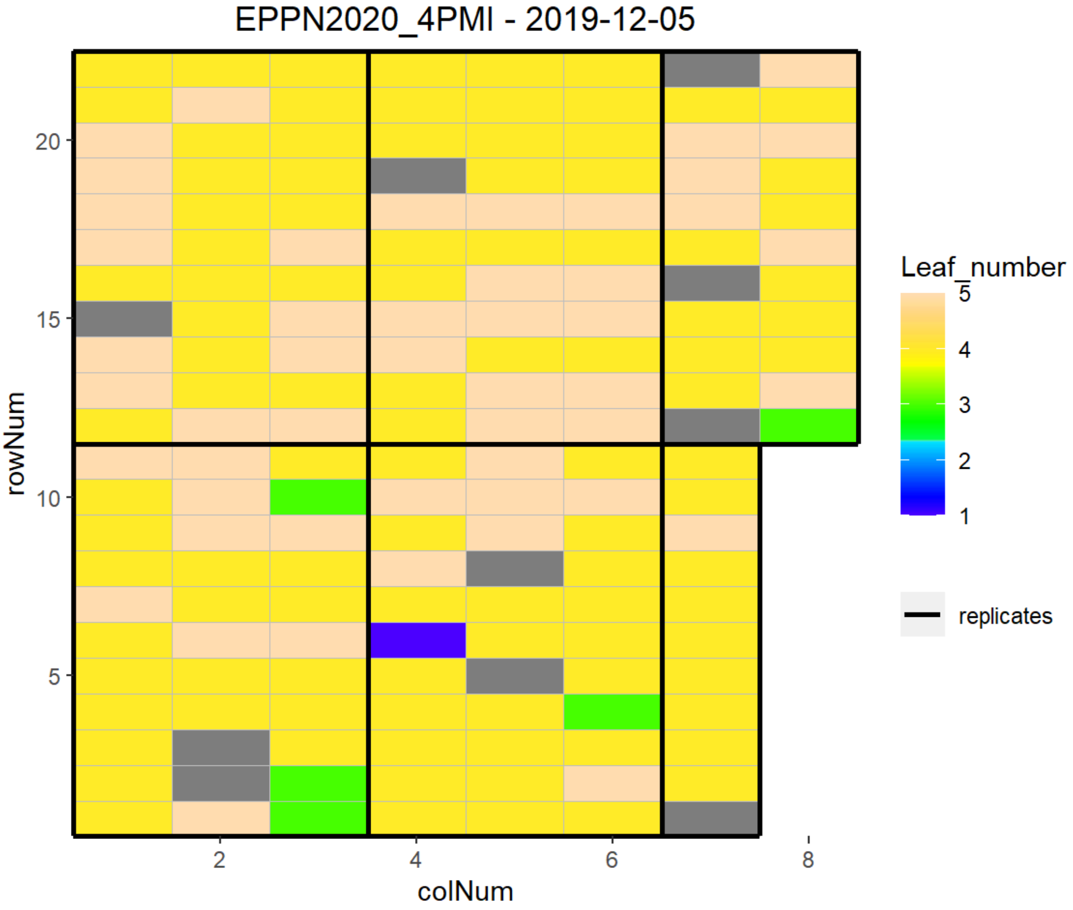

EPPN2020_4PMI - 2019-12-05
rowNum
20
15
10
5
colNum
2
4
6
8
Leaf_number
5
4
3
2
1
replicates

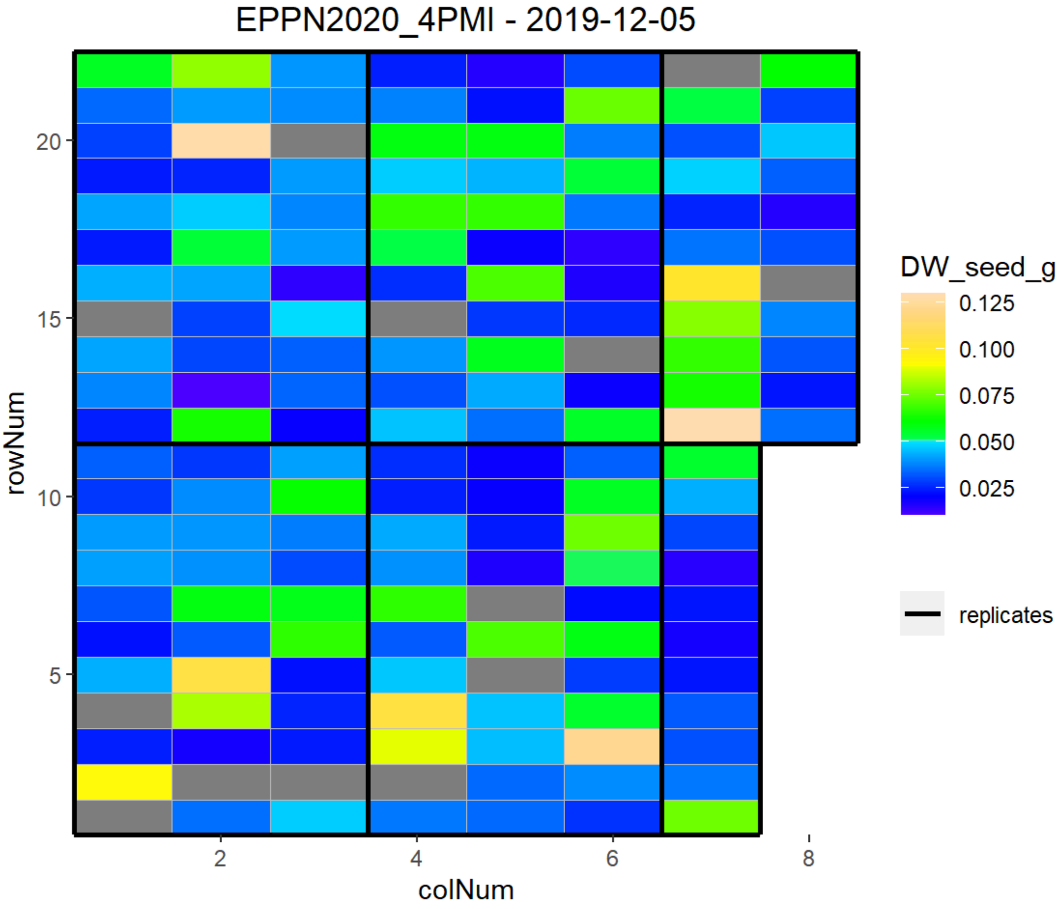

EPPN2020_4PMI - 2019-12-05
rowNum
20
15
10
5
colNum
2
4
6
8
DW_seed_g
0.15
0.10
0.05
replicates

```r
for (trait_name in traits) {
  plot(timePoint_endpoint_clean,
     plotType = "layout",
     timePoints = 1,
     traits = trait_name)
}
```


EPPN2020_4PMI - 2019-12-05
rowNum
20
15
10
5
colNum
2
4
6
8
DW_shoot_g
0.4
0.3
0.2
0.1
replicates


EPPN2020_4PMI - 2019-12-05
rowNum
20
15
10
5
colNum
2
4
6
8
DW_root_g
0.3
0.2
0.1
replicates


EPPN2020_4PMI - 2019-12-05
rowNum
20
15
10
5
colNum
2
4
6
8
Leaf_number
5
4
3
2
1
replicates


EPPN2020_4PMI - 2019-12-05
rowNum
20
15
10
5
colNum
2
4
6
8
DW_seed_g
0.125
0.100
0.075
0.050
0.025
replicates