

Eelgrass Phenology Graphics 2023

Curtis C. Bohlen and Glenn Page

2023-09-27

Contents

Introduction	1
Load Libraries	2
Explanation of Data	2
Data Properties	3
Read Data	3
Preliminary Graphics	4
Stage Graphics	4
Frequency of Mature Spathes	6
Number of Inflorescences and Spathes Per Shoot	9
Future Questions / Directions to Consider	11

Introduction

In 2023, Casco Bay Estuary Partnership (CBEP) funded a seasonal survey of the timing of flowering and seed set of eelgrass (*Zostera marina*) in Casco Bay. Scientists call the seasonal timing of flowering and other life history events “phenology”. So this study is a study of eelgrass phenology in Casco Bay, Maine.

The field project was conducted by “Team Zostera”, a growing coalition of people interested in eelgrass in the region led by Glenn Page, of SustainaMetrix.

Field observations were conducted at two eelgrass beds near Portland, Maine, one near East End Beach, in Portland, and off the north end of Mackworth Island, in Falmouth. The two sites are about 1.75 miles apart as the crow flies.

Team Zostera provided CBEP with a copy of the field data in early September of 2023. In this workbook, we assemble a “tidy” data frame based on their data and explore options for graphic presentation of the study’s results.

We anticipate little need for heavy-handed statistical methods, as a graphical summary will convey the essential seasonal information. The nature of the data also limits options for statistical modelling. Generalized linear models (on ordered data) could be useful for statistical comparison or to quantify seasonal variability, but that is not essential for a preliminary study.

A companion notebook (phenology_graphics.Rmd) provides code for and commentary about our experience importing data from Excel sheets based on the field data sheets.

Load Libraries

```
library(tidyverse)
library(readxl)
#library(emmeans) # For extracting useful "marginal" model summaries

theme_set(theme_bw())
```

Explanation of Data

Data was collected following a protocol (“SOPs”, or “Standard Operating Procedures”) developed by Jillian Carr (of the Massachusetts Bays Program) and Phil Colarusso (Of EPA’s Region 1 office, in Boston.) A copy of the protocol is available by request to the Casco Bay Estuary Partnership.

For each sampling event, a minimum of five eelgrass flowering shoots were collected. For each shoot, each side-branch (formally called a “rhipidium”) was examined, in order from the bottom of the plant (oldest) to the top of the plant (youngest).

Each side branch is a compound inflorescence, composed of multiple smaller branchlets. On each side branch (formally, a “spathe”), the flowering stage of each flowering branchlet was recorded. Order of branchlets (spathes) on the inflorescence (rhipidium) was not recorded.

Flowering stage was recorded as an ordered integer, with values from zero (spathes developed, but styles not yet appeared) to six (seeds released, shoot beginning to wither.)

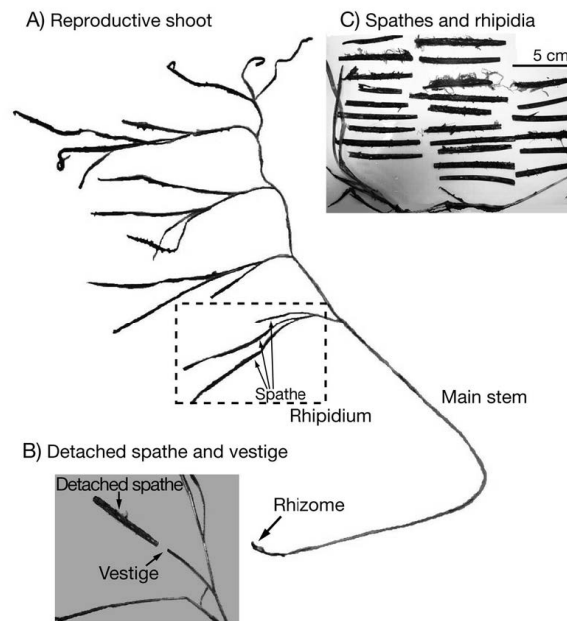


Figure 1: Eelgrass Flowering Shoot Morphology

Data Properties

The sampling protocol generates data with complex structure, including both nesting and order information that may be important.

Nesting

The nesting structure of the data includes the following:

- Spathe, nested within
- Rhodium, ordered within
- Shoot, nested within
- Date, nested within
- Site

Site and Date are not crossed, as only one site was sampled on most sampling days.

While we conducted no formal modeling in this initial analysis, it is worth reflecting on the structure of the data, and how we might treat the data in a hierarchical modelling framework.

Site and Date are naturally explanatory variables, best modeled as fixed effects. Shoots, on the other hand, are representative of all possible shoots within each site, so Shoots are best modeled as random effects. Since the order of the rhipidia on each shoot is meaningful, we need to model them somehow as fixed effects, but it is not immediately clear how. Since shoots may vary in age or number of rhipidia, order is not likely comparable across shoots, so order should be modeled within shoots.

Wow. That data structure will generate complicated models for such a simple observational study!

Read Data

See the companion workbook for details on data import process. Here we just import the previously prepared data that we saved in a “.csv” file format for convenient access.

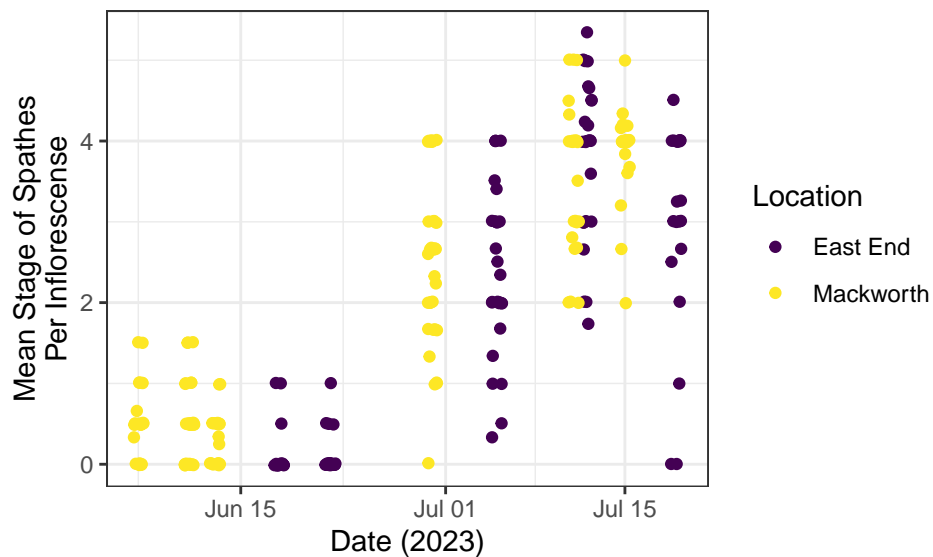
```
subfolder <- 'Working Data'
fn <- "phenology_data_2023.csv"
path <- file.path(subfolder, fn)
the_data <- read_delim(path)
#> Rows: 673 Columns: 7
#> -- Column specification -----
#> Delimiter: ","
#> chr  (3): location, Sample, Spathe
#> dbl  (3): Shoot, Rhipidium, Spathe_num
#> date (1): the_date
#>
#> i Use `spec()` to retrieve the full column specification for this data.
#> i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Preliminary Graphics

Stage Graphics

Mean Stage of Inflorescences

```
the_data %>%  
  group_by(location, the_date, Shoot, Rhipidium) %>%  
  summarize(mean_stage = mean(Spathe_num, na.rm = TRUE),  
            .groups = 'drop') %>%  
  
  ggplot(aes(the_date, mean_stage, color = location)) +  
  geom_jitter() +  
  #geom_smooth(method = "gam", formula = y~s(x, k = 3)) +  
  labs(y = "Mean Stage of Spathes\nPer Inflorescence") +  
  xlab("Date (2023)") +  
  labs(color = "Location") +  
  scale_colour_viridis_d()  
#> Warning: Removed 2 rows containing missing values (`geom_point()`).
```



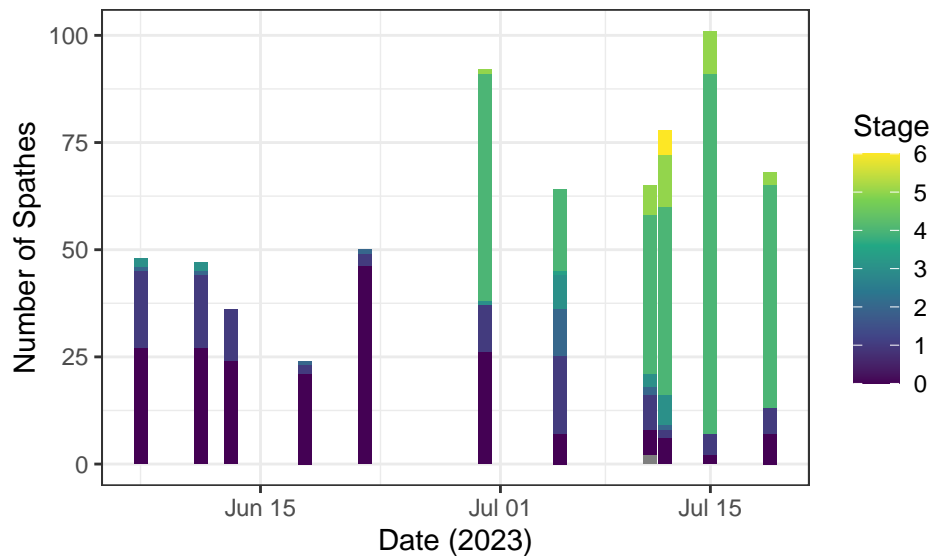
```
ggsave('Figures/mean_stage.png',  
       width = 5, height = 3)  
#> Warning: Removed 2 rows containing missing values (`geom_point()`).
```

It is not at all surprising to see the mean stage of the spathes observed climbing over the course of the season.

Bar Charts of Stage

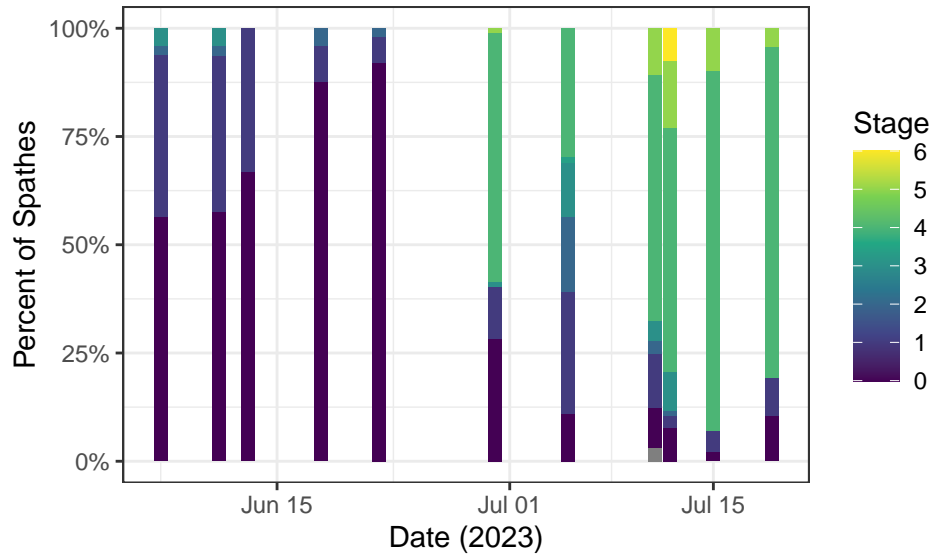
```
the_data %>%  
  ggplot(aes(x = the_date, fill = Spathe_num, group = -Spathe_num)) +
```

```
geom_bar() +
  #facet_wrap(~location) +
  ylab("Number of Spathes") +
  xlab("Date (2023)") +
  labs(fill = "Stage") +
  scale_fill_viridis_c()
```



```
ggsave('Figures/stage_bar.png',
  width = 5, height = 3)
```

```
the_data %>%
  ggplot(aes(x = the_date, fill = Spathe_num, group = -Spathe_num)) +
  geom_bar(position = "fill") +
  #facet_wrap(~location) +
  ylab("Percent of Spathes") +
  xlab("Date (2023)") +
  labs(fill = "Stage") +
  scale_y_continuous(labels = scales::percent) +
  scale_fill_viridis_c()
```



```
ggsave('Figures/stage_bar_percent.png',
       width = 5, height = 3)
```

We start to see substantial numbers and a significant percentage of spathes in Stages four or five beginning in late June or July. Both percentage and numbers appear to increase as the season progresses. It would be interesting to extend the sampling season next year to gain understanding of the end of the season as well.

Frequency of Mature Spathes

We defined spathes in stages four or five as “mature” spathes. The idea is those spathes are mature enough to potentially provide seeds when harvested.

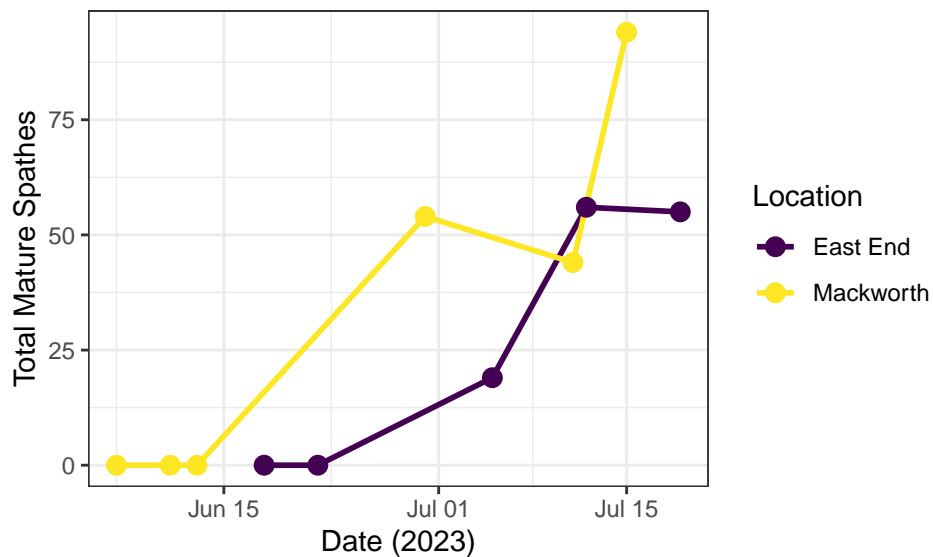
Total Number of Mature Spathes

```
the_data %>%
  mutate(is_four_five = Spathe_num == 5 | Spathe_num == 4) %>%
  group_by(location, the_date) %>%
  summarize(num_four_five = sum(is_four_five, na.rm = TRUE),
            .groups = 'drop') %>%

  ggplot(aes(x = the_date, y = num_four_five, color = location)) +
  geom_line(size = 1) +
  geom_point(size = 3) +
  #facet_wrap(~location) +
  ylab("Total Mature Spathes") +
  xlab("Date (2023)") +
  labs(color = "Location") +
  scale_colour_viridis_d()

#> Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
#> i Please use `linewidth` instead.
#> This warning is displayed once every 8 hours.
```

```
#> Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
#> generated.
```



```
ggsave('Figures/spathes_in_45.png',
       width = 5, height = 3)
```

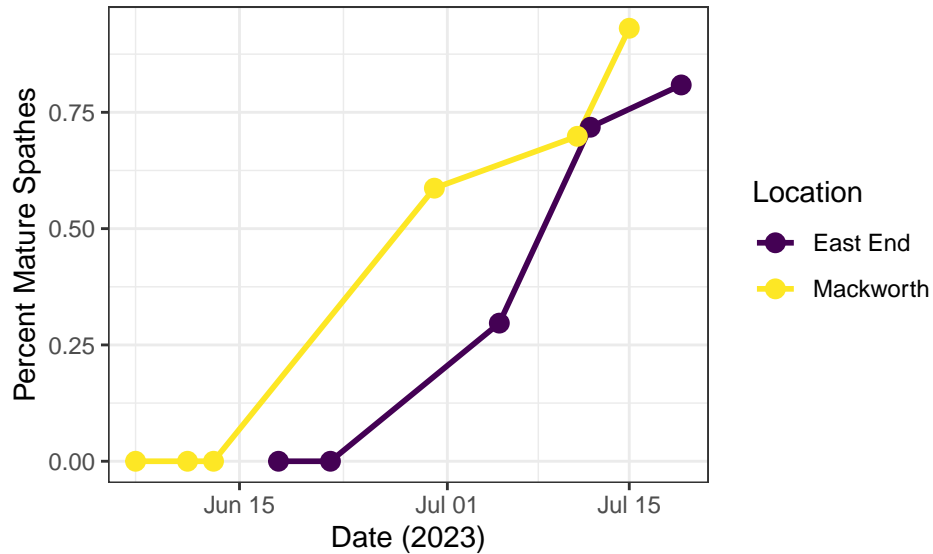
From a practical perspective, the total number of spathes that are at a stage where they might be harvested to collect seed is likely to be important for efficiency of harvest. It does not look like we sampled late enough in the season in 2023 to document a drop-off in availability of suitable spathes.

Overall Percent of Mature Spathes

This is an alternative graphic making more or less the same points.

```
the_data %>%
  filter( ! is.na(Spathe_num)) %>%
  mutate(is_four_five = Spathe_num == 5 | Spathe_num == 4) %>%
  group_by(location, the_date) %>%
  summarize(pct_four_five = sum(is_four_five)/length(is_four_five),
            .groups = 'drop') %>%

  ggplot(aes(x = the_date, y = pct_four_five, color = location)) +
  geom_line(size = 1) +
  geom_point(size = 3) +
  #facet_wrap(~location) +
  ylab("Percent Mature Spathes") +
  xlab("Date (2023)") +
  labs(color = "Location") +
  scale_colour_viridis_d()
```



```
ggsave('Figures/overall_percent_stage_45.png',
       width = 5, height = 3)
```

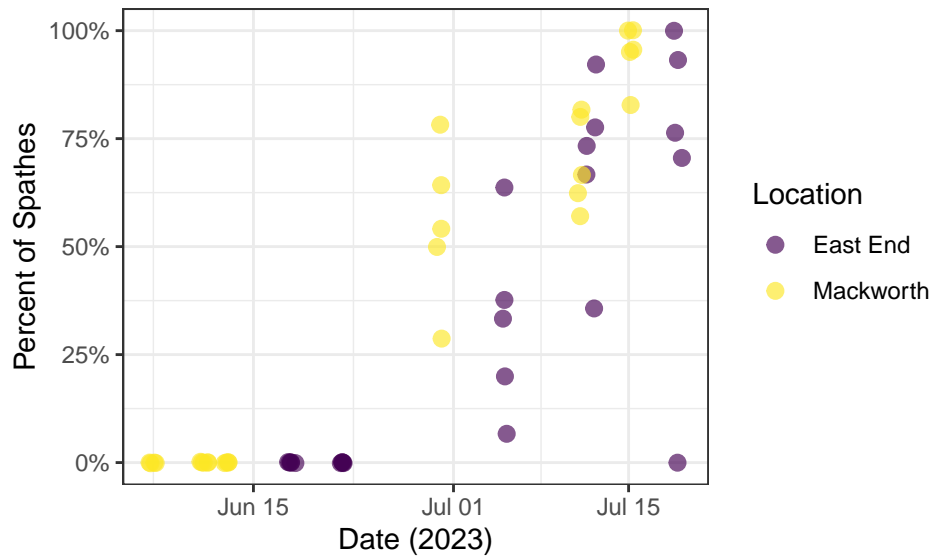
Shoot by Shoot Percent of Mature Spathes

Here the points have been moved around slightly and made partially transparent to make it clear there's a lot of shoots in June with no mature spathes.

```
the_data %>%
mutate(is_four_five = Spathe_num == 5 | Spathe_num == 4) %>%
group_by(location, the_date, Shoot) %>%
summarize(pct_45 = sum(is_four_five, na.rm = TRUE)/
           (length(is_four_five) - sum(is.na(is_four_five))),
           .groups = 'drop') %>%

ggplot(aes(the_date, pct_45, color = location)) +
  #geom_point(size = 2.5, alpha = 0.5) +
  geom_jitter(size = 2.5, alpha = 0.65) +

  #geom_smooth(method = 'gam', formula = y~s(x, k = 3)) +
  ylab("Percent of Spathes") +
  xlab("Date (2023)") +
  labs(color = "Location") +
  scale_y_continuous(labels = scales::percent) +
  scale_color_viridis_d()
```

```
ggsave('Figures/Shoot_percent_stage_45.png',
       width = 5, height = 3)
```

This offers a bit more understanding of variability among shoots. By mid-July, most shoots have at least 50% mature spathes, suggesting that's likely to be a good time to harvest. But the proportion of spathes that are mature varies quite a bit.

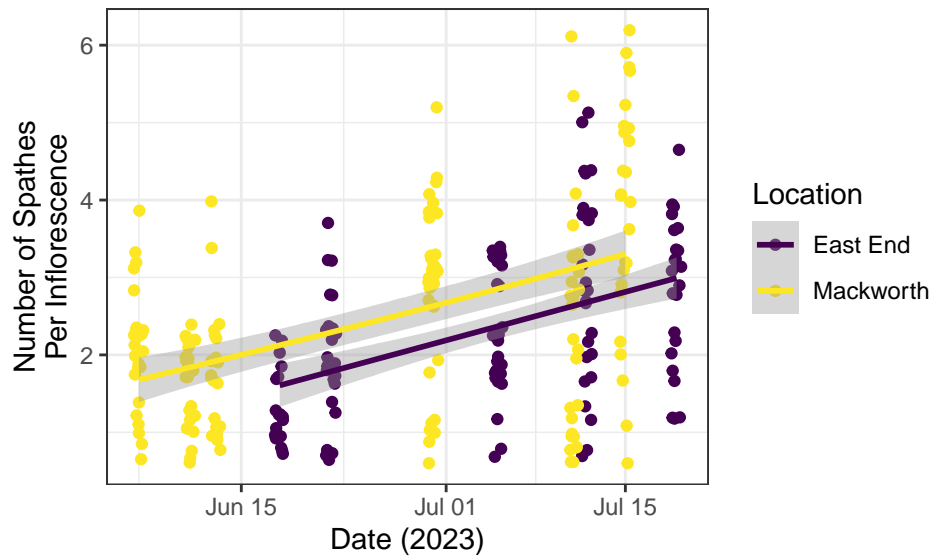
As before, the data does not extend late enough in the season to show a drop off in the number of mature spathes.

Number of Inflorescences and Spathes Per Shoot

Number of Spathes per Inflorescence

If there are strong seasonal patterns in the number of spathes per inflorescence, or the number of inflorescence per shoot, that might also influence the ideal time for harvesting seed. It is going to be best to sample when we can get the most mature spathes, not the highest percentage of mature spathes.

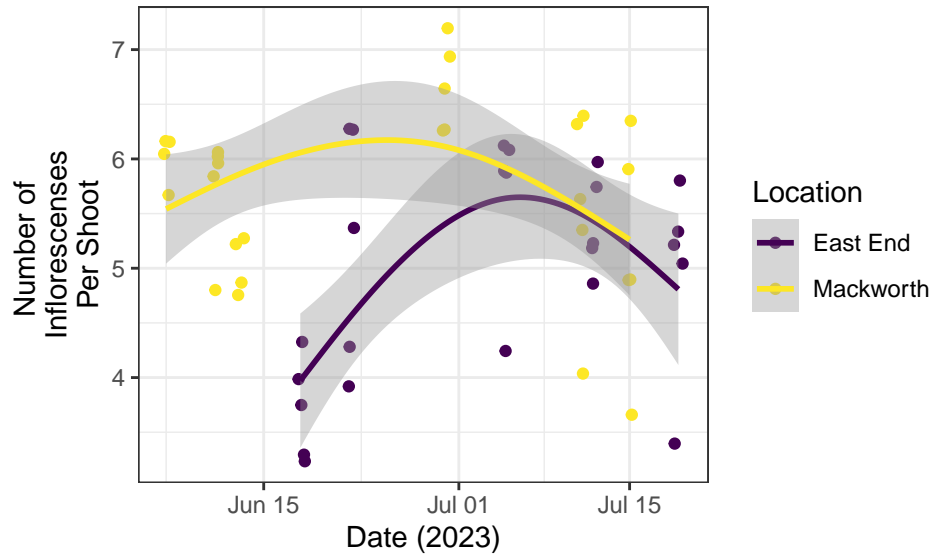
```
the_data %>%
  group_by(location, the_date, Shoot, Rhipidium) %>%
  summarize(num_spathes = n(),
            .groups = 'drop') %>%
  ggplot(aes(the_date, num_spathes, color = location)) +
  geom_jitter() +
  geom_smooth(method = "gam", formula = y~s(x, k = 3)) +
  ylab("Number of Spathes\nPer Inflorescence") +
  xlab("Date (2023)") +
  labs(color = "Location") +
  scale_colour_viridis_d()
```



```
ggsave('Figures/spathes_per_inf.png',
       width = 5, height = 3)
```

Number of Inflorescences Per Shoot

```
the_data %>%
  group_by(location, the_date, Shoot, Rhipidium) %>%
  summarize(.groups = 'drop_last') %>%
  summarize(num_Rhipidia = n(),
            .groups = 'drop') %>%
  ggplot(aes(the_date, num_Rhipidia, color = location)) +
  geom_jitter() +
  geom_smooth(method = "gam", formula = y~s(x, k = 3)) +
  ylab("Number of \nInflorescences\nPer Shoot") +
  xlab("Date (2023)") +
  labs(color = "Location") +
  scale_colour_viridis_d()
```



```
ggsave('Figures/infl_per_shoot.png',
       width = 5, height = 3)
```

The number of spathes per inflorescence continued to increase through the summer, while the number of inflorescences per shoot appears to peak sometime in late June or early July. Given the limited data, those apparent patterns may reflect sampling variability (note the wide error bands). This deserves additional attention in future years.

Future Questions / Directions to Consider

1. Differences in Stage by order of Inflorescence on the Shoot.
2. Hierarchical Generalized Linear Models of Stage, based on ordered data.