

WK3

2025-07-02

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.2      v tibble    3.2.1
## v lubridate  1.9.4      v tidyr     1.3.1
## v purrr      1.0.4
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
library(lubridate)
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test
```

```
library(purrr)
library(readr)
library(ggthemes)
#setwd("C:/Users/Tobyz/Desktop/S. tort-light-growth/Data")
```

```
import plant data
```

```
plant <- read_csv("Data/WL2-2023_Size_Combined.csv") %>%
  clean_names() %>%
  mutate(survey_date = as.Date(survey_date, format = "%m/%d/%Y"))
summary(plant)
```

```
##   survey_date      block      genotype      pop_mf
## Min.   :2023-07-03 Length:17336 Length:17336 Length:17336
## 1st Qu.:2023-08-02 Class :character Class :character Class :character
## Median :2023-08-30 Mode  :character Mode  :character Mode  :character
## Mean   :2023-08-28
## 3rd Qu.:2023-09-20
```

```
## Max.      :2023-10-20
##
## parent_pop      mf      rep      height_cm
## Length:17336    Min.    : 1.000    Min.    : 1.000    Min.    : 0.100
## Class :character 1st Qu.: 2.000    1st Qu.: 4.000    1st Qu.: 1.700
## Mode  :character Median : 5.000    Median : 8.000    Median : 3.100
##                      Mean   : 4.584    Mean   : 7.932    Mean   : 4.491
##                      3rd Qu.: 6.000    3rd Qu.:11.000    3rd Qu.: 5.700
##                      Max.    :14.000    Max.    :27.000    Max.    :39.400
##                      NA's    :8762
## long_leaf_cm    survey_notes
## Min.      :0.100    Length:17336
## 1st Qu.:1.600    Class :character
## Median :2.500    Mode  :character
## Mean     :2.599
## 3rd Qu.:3.500
## Max.     :9.000
## NA's     :9350
```

consolidate light measurement to a weekly measurement

```
#import light data
light_raw <- read_csv("Data/IntBioHalfHourTable_clean.txt")
```

```
## Rows: 4063 Columns: 139
## -- Column specification -----
## Delimiter: ","
## dbl  (138): RECORD, BattV_Max, PTemp_C_Max, SlrW_Avg, SlrW_Max, SlrW_Min, Sl...
## dtm   (1): TIMESTAMP
##
## 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.
```

```
#weekly measurement
weekly_light <- light_raw %>%
  mutate(
    timestamp = ymd_hms(TIMESTAMP),
    SlrW_Avg = as.numeric(SlrW_Avg), # turn into number format
    week = floor_date(timestamp, "week")
  ) %>%
  group_by(week) %>%
  summarise(
    weekly_avg_SlrW = mean(SlrW_Avg, na.rm = TRUE),
    .groups = "drop"
  )
```

```
## Warning: There was 1 warning in 'mutate()'.
## i In argument: 'timestamp = ymd_hms(TIMESTAMP)'.
## Caused by warning:
## ! 84 failed to parse.
```

```
# result
print(weekly_light)
```

```
## # A tibble: 14 x 2
##   week          weekly_avg_SlrW
##   <dtm>          <dbl>
## 1 2023-07-30 00:00:00      280.
## 2 2023-08-06 00:00:00      277.
## 3 2023-08-13 00:00:00      186.
## 4 2023-08-20 00:00:00      184.
## 5 2023-08-27 00:00:00      200.
## 6 2023-09-03 00:00:00      211.
## 7 2023-09-10 00:00:00      204.
## 8 2023-09-17 00:00:00      189.
## 9 2023-09-24 00:00:00      159.
## 10 2023-10-01 00:00:00      138.
## 11 2023-10-08 00:00:00      133.
## 12 2023-10-15 00:00:00      135.
## 13 2023-10-22 00:00:00      116.
## 14 NA              -0.616
```

compute growth from size measurements

```
plant_growth <- plant %>%
  unite("PID", genotype:rep, sep = "_") %>%
  mutate(survey_date = as.Date(survey_date)) %>%
  select(PID, survey_date, height_cm)

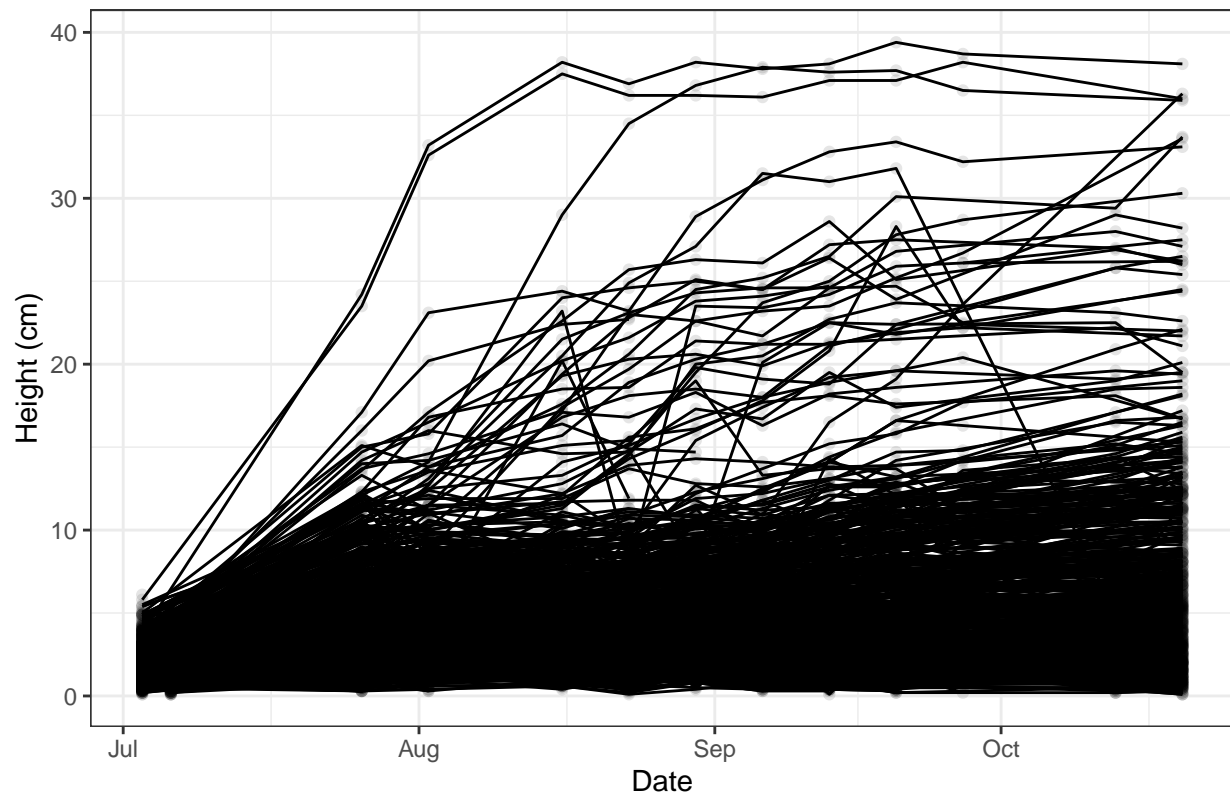
# Will write this data once in a csv and read it in next time
#write_csv(plant_growth, "Data/plant_growth.csv")
#plant_growth = read_csv("Data/plant_growth.csv")

#plot growth curve
ggplot(plant_growth, aes(x = survey_date, y = height_cm, group = PID)) +
  geom_line() +
  geom_point(alpha = 0.1) +
  labs(title = "Growth Curve of Plants", x = "Date", y = "Height (cm)") +
  theme_bw() +
  theme(legend.position = "bottom")
```

```
## Warning: Removed 8715 rows containing missing values or values outside the scale range
## ('geom_line()').
```

```
## Warning: Removed 8762 rows containing missing values or values outside the scale range
## ('geom_point()').
```

Growth Curve of Plants



```
#check data on these two dates
plant_growth %>%
  filter(survey_date %in% as.Date(c("2023-09-27", "2023-10-13"))) %>%
  group_by(PID, survey_date) %>%
  summarise(has_data = any(!is.na(height_cm)), .groups = "drop") %>%
  pivot_wider(names_from = survey_date, values_from = has_data)
```

```
## # A tibble: 1,573 x 3
##   PID          '2023-10-13' '2023-09-27'
##   <chr>          <lgl>      <lgl>
## 1 BH_1_10_BH_1_BH_1_10 TRUE      NA
## 2 BH_1_11_BH_1_BH_1_11 FALSE     NA
## 3 BH_1_12_BH_1_BH_1_12 TRUE      NA
## 4 BH_1_13_BH_1_BH_1_13 TRUE      NA
## 5 BH_1_1_BH_1_BH_1_1  TRUE      NA
## 6 BH_1_2_BH_1_BH_1_2  FALSE     NA
## 7 BH_1_3_BH_1_BH_1_3  NA        FALSE
## 8 BH_1_4_BH_1_BH_1_4  NA        TRUE
## 9 BH_1_5_BH_1_BH_1_5  NA        FALSE
## 10 BH_1_6_BH_1_BH_1_6  NA        FALSE
## # i 1,563 more rows
```

```
plant_growth %>%
  filter(survey_date %in% as.Date(c("2023-09-27", "2023-10-13"))) %>%
  group_by(PID, survey_date) %>%
```

```

summarise(has_data = any(!is.na(height_cm)), .groups = "drop") %>%
pivot_wider(names_from = survey_date, values_from = has_data) %>%
mutate(
  status = case_when(
    `2023-09-27` == TRUE & `2023-10-13` == TRUE ~ "both",
    `2023-09-27` == TRUE & (`2023-10-13` == FALSE | is.na(`2023-10-13`)) ~ "only 9/27",
    (`2023-09-27` == FALSE | is.na(`2023-09-27`)) & `2023-10-13` == TRUE ~ "only 10/13",
    TRUE ~ "neither"
  )
) %>%
count(status)

```

```

## # A tibble: 3 x 2
##   status      n
##   <chr>    <int>
## 1 neither    1011
## 2 only 10/13   336
## 3 only 9/27   226

```

Investigate or filter out plants that show negative growth

```

#find out plants with negative growth
plant_growth %>%
  arrange(PID, survey_date) %>% # arrange in time sequence
  group_by(PID) %>%
  mutate(growth = height_cm - lag(height_cm)) %>% # find out the diff btw nearby dates
  summarise(has_negative_growth = any(growth < 0, na.rm = TRUE)) %>%
  filter(has_negative_growth) -> neg_growth_plants
neg_growth_plants

```

```

## # A tibble: 826 x 2
##   PID                has_negative_growth
##   <chr>              <lgl>
## 1 BH_1_10_BH_1_BH_1_10 TRUE
## 2 BH_1_12_BH_1_BH_1_12 TRUE
## 3 BH_1_13_BH_1_BH_1_13 TRUE
## 4 BH_1_1_BH_1_BH_1_1 TRUE
## 5 BH_1_4_BH_1_BH_1_4 TRUE
## 6 BH_1_7_BH_1_BH_1_7 TRUE
## 7 BH_2_10_BH_2_BH_2_10 TRUE
## 8 BH_2_11_BH_2_BH_2_11 TRUE
## 9 BH_2_12_BH_2_BH_2_12 TRUE
## 10 BH_2_13_BH_2_BH_2_13 TRUE
## # i 816 more rows

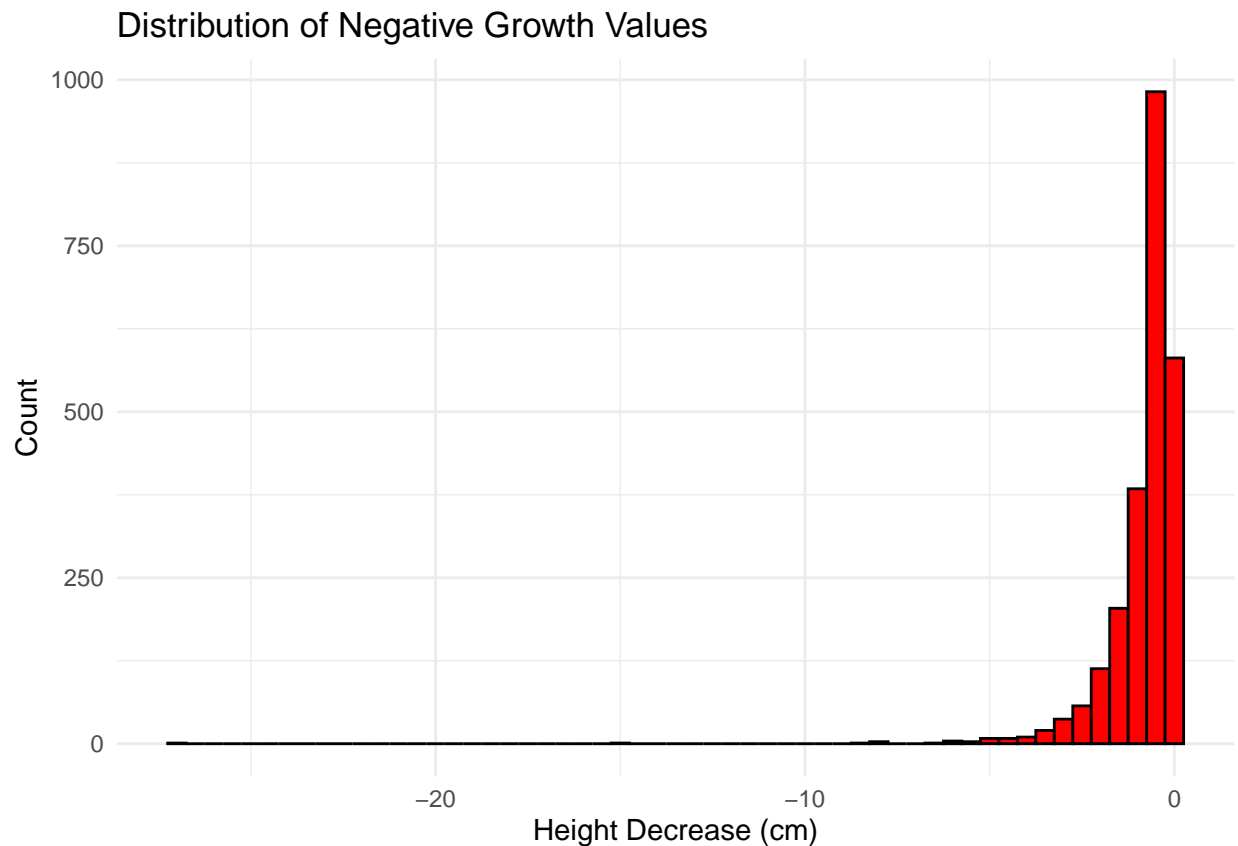
```

```

#find out tolerance value
neg_growth_values <- plant_growth %>%
  arrange(PID, survey_date) %>%
  group_by(PID) %>%
  mutate(growth = height_cm - lag(height_cm)) %>%
  ungroup() %>%
  filter(growth < 0)

```

```
ggplot(neg_growth_values, aes(x = growth)) +
  geom_histogram(binwidth = 0.5, fill = "red", color = "black") +
  labs(
    title = "Distribution of Negative Growth Values",
    x = "Height Decrease (cm)",
    y = "Count"
  ) +
  theme_minimal()
```



```
#filter out plants with negative growth < -5
plant_growth_cleaned <- plant_growth

repeat {
  plant_growth_cleaned <- plant_growth_cleaned %>%
    arrange(PID, survey_date) %>%
    group_by(PID) %>%
    mutate(growth = height_cm - lag(height_cm)) %>%
    filter(is.na(growth) | growth >= -5) %>%
    select(-growth) %>%
    ungroup()

  check <- plant_growth_cleaned %>%
    arrange(PID, survey_date) %>%
    group_by(PID) %>%
    mutate(growth = height_cm - lag(height_cm)) %>%

```

```

    filter(growth < -5)

    if (nrow(check) == 0) break
  }

  #plot growth curve
  ggplot(plant_growth_cleaned, aes(x = survey_date, y = height_cm, group = PID)) +
    geom_line() +
    geom_point(alpha = 0.1) +
    labs(title = "Growth Curve of Plants without Negative Growth < -5", x = "Date", y = "Height (cm)") +
    theme_bw() +
    theme(legend.position = "bottom")

```

```

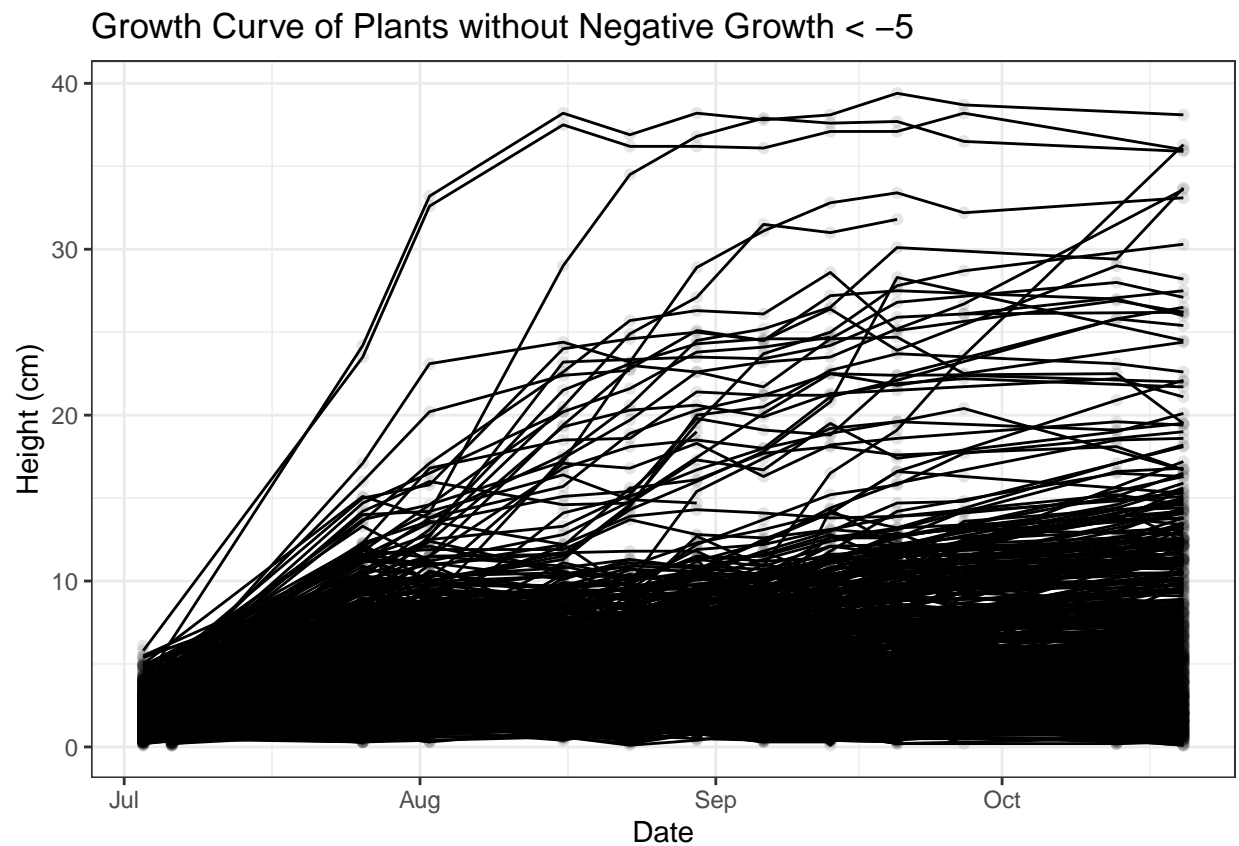
## Warning: Removed 8715 rows containing missing values or values outside the scale range
## ('geom_line()').

```

```

## Warning: Removed 8762 rows containing missing values or values outside the scale range
## ('geom_point()').

```



measure growth via relative growth

```

plant_growth_rel <- plant_growth_cleaned %>%
  arrange(PID, survey_date) %>%
  group_by(PID) %>%
  mutate(

```

```

    prev_height = lag(height_cm),
    growth = height_cm - prev_height,
    relative_growth = growth / prev_height
  ) %>%
  ungroup()
plant_growth_rel

```

```
## # A tibble: 17,312 x 6
```

```

##   PID          survey_date height_cm prev_height growth relative_growth
##   <chr>          <date>      <dbl>      <dbl>  <dbl>      <dbl>
## 1 BH_1_10_BH_1_BH_1_10 2023-07-03      2.5        NA    NA         NA
## 2 BH_1_10_BH_1_BH_1_10 2023-07-26      3.1        2.5    0.6        0.24
## 3 BH_1_10_BH_1_BH_1_10 2023-08-02      3.9        3.1    0.8        0.258
## 4 BH_1_10_BH_1_BH_1_10 2023-08-16      2.7        3.9   -1.2       -0.308
## 5 BH_1_10_BH_1_BH_1_10 2023-08-23      3.1        2.7    0.4        0.148
## 6 BH_1_10_BH_1_BH_1_10 2023-08-30      3.1        3.1    0         0
## 7 BH_1_10_BH_1_BH_1_10 2023-09-06      4          3.1    0.9        0.290
## 8 BH_1_10_BH_1_BH_1_10 2023-09-13      5.9        4       1.9        0.475
## 9 BH_1_10_BH_1_BH_1_10 2023-09-20      4.6        5.9   -1.3       -0.220
## 10 BH_1_10_BH_1_BH_1_10 2023-10-13      4.3        4.6  -0.300     -0.0652
## # i 17,302 more rows

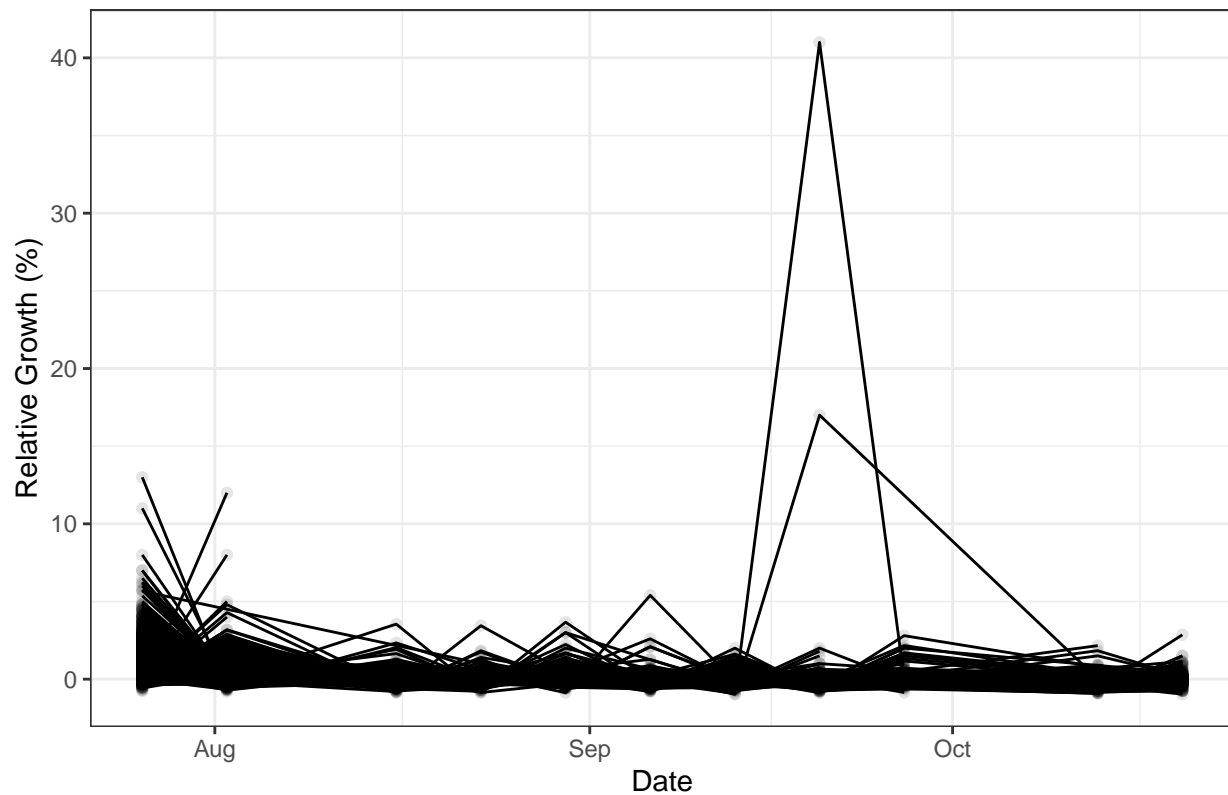
```

```

plant_growth_rel%>%
  filter(!is.na(relative_growth)) %>%
  ggplot(aes(survey_date, relative_growth, group= PID))+
  geom_line()+
  geom_point(alpha=0.1)+
  labs(title="Relative Growth Curve of Plants without Negative Growth < -5", x = "Date", y = "Relative Growth")
  theme_bw()

```


Relative Growth Curve of Plants without Negative Growth < -5

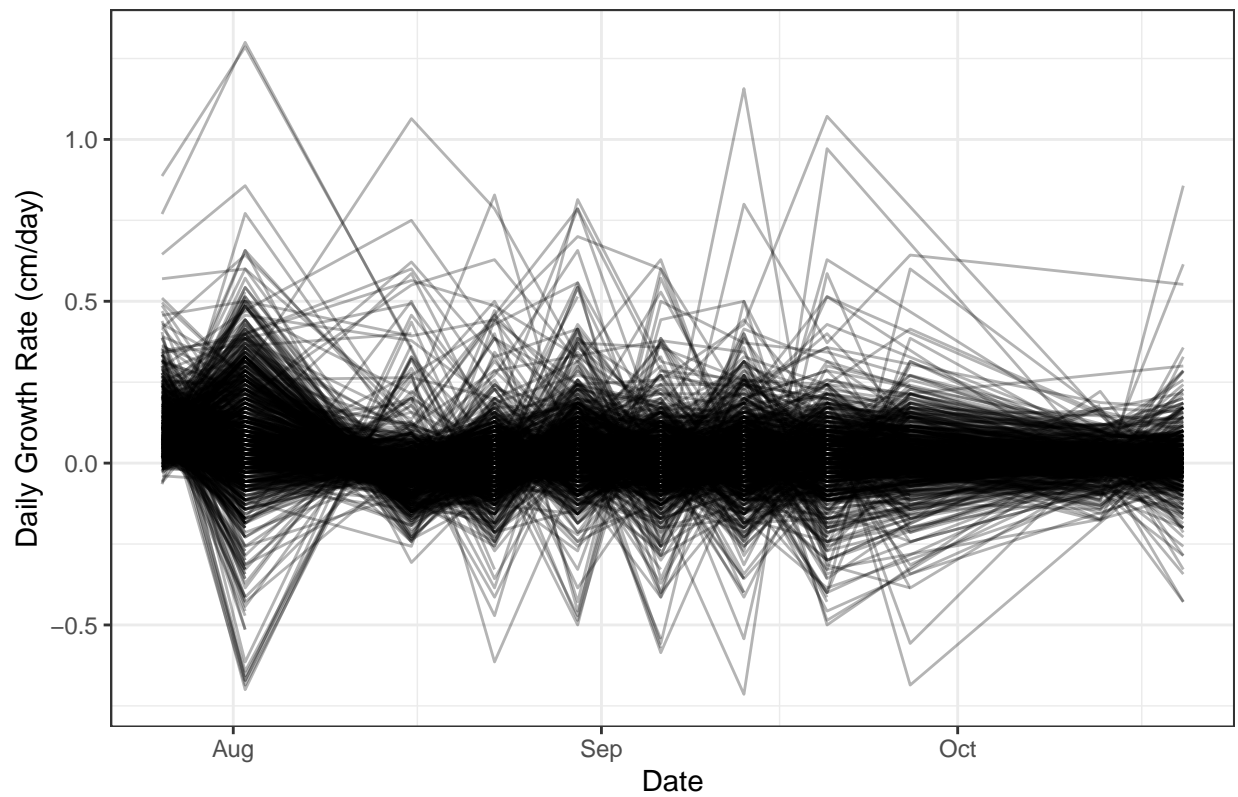


Measure Growth via Daily Growth Rate

```
#define daily growth rate
plant_growth_daily <- plant_growth_cleaned %>%
  arrange(PID, survey_date) %>%
  group_by(PID) %>%
  mutate(
    prev_height = lag(height_cm),
    prev_date = lag(survey_date),
    days_elapsed = as.numeric(survey_date - prev_date),
    daily_growth = (height_cm - prev_height) / days_elapsed
  ) %>%
  ungroup()

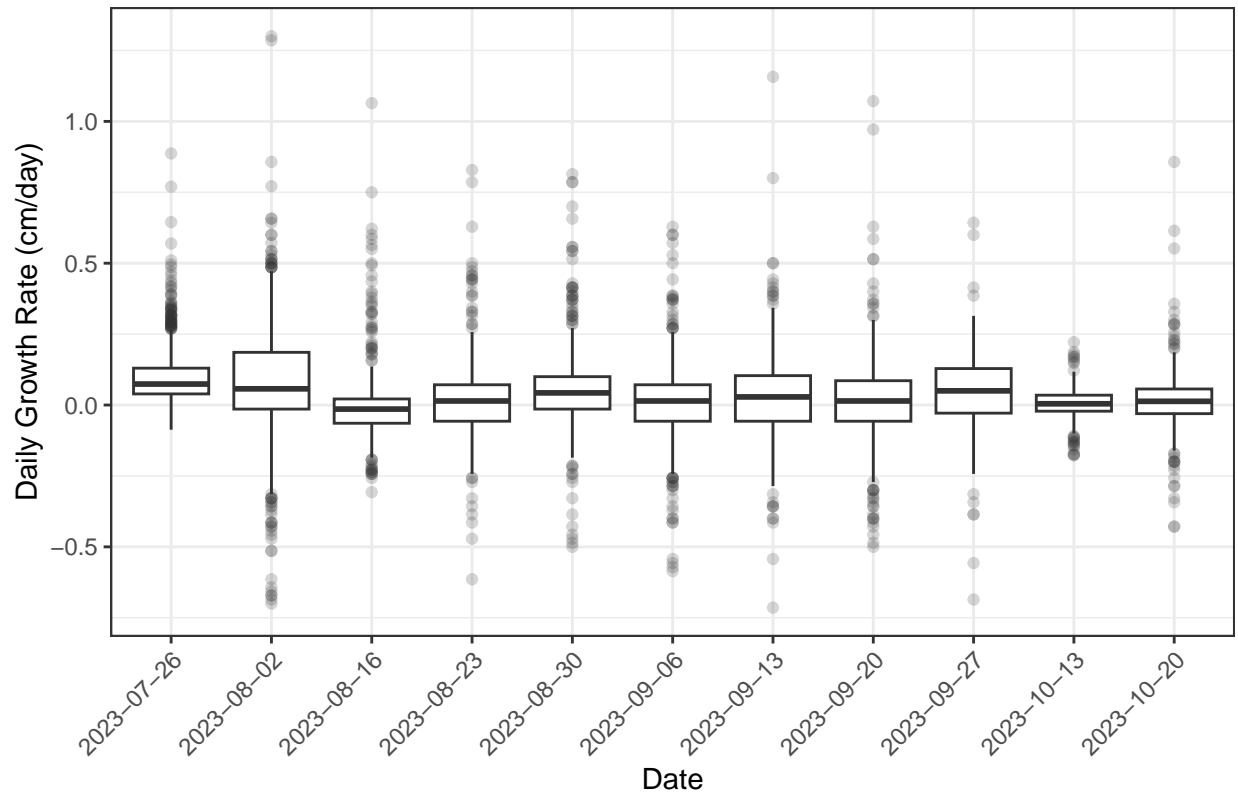
#Line graph
plant_growth_daily %>%
  filter(!is.na(daily_growth)) %>%
  ggplot(aes(x = survey_date, y = daily_growth, group = PID)) +
  geom_line(alpha = 0.3) +
  labs(title = "Daily Growth Rate of Plants without Negative Growth < -5", y = "Daily Growth Rate (cm/d)") +
  theme_bw()
```

Daily Growth Rate of Plants without Negative Growth < -5



```
#boxplot
plant_growth_daily %>%
  filter(!is.na(daily_growth), days_elapsed > 0) %>%
  mutate(date_factor = factor(survey_date, levels = sort(unique(survey_date)))) %>%
  ggplot(aes(x = date_factor, y = daily_growth)) +
  geom_boxplot(outlier.alpha = 0.2) +
  labs(title = "Distribution of Daily Growth Rates without Negative Growth < -5", x = "Date", y = "Daily Growth Rate (cm/day)") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Distribution of Daily Growth Rates without Negative Growth < -5



Correlate Growth with Solar Radiation

```
#Align plant growth data to week
plant_weekly <- plant_growth_daily %>%
  filter(!is.na(daily_growth), days_elapsed > 0) %>%
  mutate(week = floor_date(survey_date, "week"))

#Adds `weekly_avg_SlrW` to plant data
plant_with_light <- plant_weekly %>%
  left_join(weekly_light, by = "week")

#Calculate correlation
cor_result <- cor(
  plant_with_light$daily_growth,
  plant_with_light$weekly_avg_SlrW,
  use = "complete.obs"
)

print(cor_result)
```

```
## [1] 0.1299395
```

```
ggplot(plant_with_light, aes(x = weekly_avg_SlrW, y = daily_growth)) +
  geom_point(alpha = 0.2) +
  geom_smooth(method = "lm", color = "red") +
```

```
labs(title = "Correlation between Light and Growth", x = "Weekly Avg Light (SlrW)", y = "Daily Growth  
theme_bw()
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 1069 rows containing non-finite outside the scale range  
## ('stat_smooth()').
```

```
## Warning: Removed 1069 rows containing missing values or values outside the scale range  
## ('geom_point()').
```

