Preliminary Data Analysis

Project Description:

The data analysis will be conducted to evaluate the effect of zucchini planting date on the need for pollinator supplementation with managed bee hives. The data being used is unpublished and represents proportion analysis of bee visits over time. Dr. Ashley Leach, as co-PI, has been conducting experiments established across three field sites (Wooster, Willard, and Piketon). For this preliminary analysis, data from the Wooster site is being used, with additional site data to be incorporated in future analyses as it becomes available.

Predictor (Categorical) Variables:

1. Planting Date (June 1 vs July 1) - proxy for bloom time

2. Pollination Treatment (With managed bumblebees vs Without supplementation)

Response (Continuous) Variables:

1. Average number of fruit per plant (yield measure)

2. Average number of overall visits per flower per minute (pollination measure)

Loading the dataset into R:

```
rm(list=ls()) #clearing R's Brain

library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✓ dplyr     1.1.4     ✓ readr     2.1.5
✓ forcats   1.0.0     ✓ stringr   1.5.2
✓ ggplot2   4.0.0     ✓ tibble    3.3.0
✓ lubridate 1.9.4     ✓ tidyr     1.3.1
✓ purrr     1.1.0
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
errors
```

```
library(dplyr)
library(readxl)
library(ggplot2)

squash_bees_df <- read_excel("C:/Users/Rosita/OneDrive - The Ohio State University/Documents/AU
         2025/Ent Tech&DataAnalysis/R_dataset1/Preliminary_data/pollinator_records.xlsx",
         sheet="Pollinator.Records", "A1:Q97" )
```

```
summary(squash_bees_df)
```

```
     date                       treatment            planting    
 Min.   :2025-07-12 00:00:00   Length:96          Min.   :1.00  
 1st Qu.:2025-07-14 06:00:00   Class :character   1st Qu.:1.00  
 Median :2025-07-18 12:00:00   Mode  :character   Median :1.00  
 Mean   :2025-07-21 12:00:00                      Mean   :1.25  
 3rd Qu.:2025-07-25 18:00:00                      3rd Qu.:1.25  
 Max.   :2025-08-06 00:00:00                      Max.   :2.00  
    site            observation       sampling     squash.bees    
 Length:96          Min.   : 1.00   Min.   :1.00   Min.   : 0.000  
 Class :character   1st Qu.: 3.75   1st Qu.:1.00   1st Qu.: 3.000  
 Mode  :character   Median : 6.50   Median :1.50   Median : 6.000  
                    Mean   : 6.50   Mean   :1.75   Mean   : 6.177  
                    3rd Qu.: 9.25   3rd Qu.:2.25   3rd Qu.: 9.000  
                    Max.   :12.00   Max.   :3.00   Max.   :16.000  
  bumblebees       honeybees        sweat.bees         other      
 Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0  
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0  
 Median :0.0000   Median :0.0000   Median :0.0000   Median :0  
 Mean   :0.4688   Mean   :0.4375   Mean   :0.0625   Mean   :0  
 3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:0  
 Max.   :7.0000   Max.   :3.0000   Max.   :1.0000   Max.   :0  
    total           flowers         male           female      
 Min.   : 0.000   Min.   :1.0   Min.   :0.000   Min.   :0.0000  
 1st Qu.: 4.000   1st Qu.:1.0   1st Qu.:1.000   1st Qu.:0.0000  
 Median : 6.500   Median :1.0   Median :1.000   Median :0.0000  
 Mean   : 7.146   Mean   :1.5   Mean   :1.062   Mean   :0.4375  
 3rd Qu.:11.000   3rd Qu.:2.0   3rd Qu.:1.000   3rd Qu.:1.0000  
 Max.   :20.000   Max.   :4.0   Max.   :3.000   Max.   :2.0000  
 visits.flower    visits.flower.min
 Min.   : 0.000   Min.   :0.0000  
 1st Qu.: 2.500   1st Qu.:0.8333  
 Median : 5.000   Median :1.6667  
 Mean   : 5.507   Mean   :1.8356  
 3rd Qu.: 8.000   3rd Qu.:2.6667  
 Max.   :15.000   Max.   :5.0000  
```

```
head(squash_bees_df)
```

```
# A tibble: 6 × 17
  date                treatment  planting site  observation sampling squash.bees
  <dttm>              <chr>         <dbl> <chr>       <dbl>    <dbl>       <dbl>
1 2025-07-12 00:00:00 Supplemen…        1 woos…           1        1           5
2 2025-07-12 00:00:00 Supplemen…        1 woos…           2        1           1
3 2025-07-12 00:00:00 Supplemen…        1 woos…           3        1          15
4 2025-07-12 00:00:00 Supplemen…        1 woos…           4        1           5
5 2025-07-12 00:00:00 Supplemen…        1 woos…           5        1           2
6 2025-07-12 00:00:00 Supplemen…        1 woos…           6        1           0
```

```
# ℹ 10 more variables: bumblebees <dbl>, honeybees <dbl>, sweat.bees <dbl>,
#   other <dbl>, total <dbl>, flowers <dbl>, male <dbl>, female <dbl>,
#   visits.flower <dbl>, visits.flower.min <dbl>
```

```
num_data <- nrow(squash_bees_df)
print(num_data)
```

```
[1] 96
```

```
squash_bees_df$treatment <- as.factor(squash_bees_df$treatment)
str(squash_bees_df)
```

```
tibble [96 × 17] (S3: tbl_df/tbl/data.frame)
 $ date             : POSIXct[1:96], format: "2025-07-12" "2025-07-12" ...
 $ treatment        : Factor w/ 2 levels "Not supplemented",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ planting         : num [1:96] 1 1 1 1 1 1 1 1 1 1 ...
 $ site             : chr [1:96] "wooster" "wooster" "wooster" "wooster" ...
 $ observation      : num [1:96] 1 2 3 4 5 6 7 8 9 10 ...
 $ sampling         : num [1:96] 1 1 1 1 1 1 1 1 1 1 ...
 $ squash.bees      : num [1:96] 5 1 15 5 2 0 4 8 1 0 ...
 $ bumblebees       : num [1:96] 0 0 0 0 0 0 0 0 0 0 ...
 $ honeybees        : num [1:96] 0 0 0 0 0 0 0 1 0 0 ...
 $ sweat.bees       : num [1:96] 0 0 0 0 0 0 0 0 0 0 ...
 $ other            : num [1:96] 0 0 0 0 0 0 0 0 0 0 ...
 $ total            : num [1:96] 5 1 15 5 2 0 4 9 1 0 ...
 $ flowers          : num [1:96] 1 2 2 2 2 3 2 2 1 1 ...
 $ male             : num [1:96] 1 1 2 1 2 2 1 2 1 1 ...
 $ female           : num [1:96] 0 1 0 1 0 1 1 0 0 0 ...
 $ visits.flower    : num [1:96] 5 0.5 7.5 2.5 1 0 2 4.5 1 0 ...
 $ visits.flower.min: num [1:96] 1.667 0.167 2.5 0.833 0.333 ...
```

```
yield_df <- read_excel("C:/Users/Rosita/OneDrive - The Ohio State University/Documents/AU
         2025/Ent Tech&DataAnalysis/R_dataset1/Preliminary_data/yield_prelim.xlsx",
         sheet="yield_prelim", "A1:Q26")

summary(yield_df)
```

```
     year          site            treatment            planting
 Min.   :2025   Length:25          Length:25          Min.   :1.00
 1st Qu.:2025   Class :character   Class :character   1st Qu.:1.00
 Median :2025   Mode  :character   Mode  :character   Median :2.00
 Mean   :2025                                         Mean   :1.56
 3rd Qu.:2025                                         3rd Qu.:2.00
 Max.   :2025                                         Max.   :2.00
    harvest      num_harvest      damaged add    weight_harvest
 Min.   :1.00   Min.   :  6.00   Min.   : 0.00   Length:25
 1st Qu.:2.00   1st Qu.: 15.00   1st Qu.: 0.00   Class :character
 Median :2.00   Median : 45.00   Median : 0.00   Mode  :character
 Mean   :2.12   Mean   : 57.88   Mean   : 7.16
```

```
 3rd Qu.:3.00   3rd Qu.: 61.00   3rd Qu.: 0.00
 Max.   :3.00   Max.   :218.00   Max.   :61.00
   Cull_poll     Cull_poll_weight     Cull_other   Cull_other_weight
 Min.   : 0.00   Length:25          Min.   :0.00   Length:25
 1st Qu.: 0.00   Class :character   1st Qu.:0.00   Class :character
 Median : 1.00   Mode  :character   Median :0.00   Mode  :character
 Mean   : 2.44                      Mean   :0.44
 3rd Qu.: 3.00                      3rd Qu.:0.00
 Max.   :21.00                      Max.   :3.00
     stand       functional stand  yield.plant     yield.plant_weight
 Min.   :32.00   Min.   :20.00    Min.   :0.1034   Length:25
 1st Qu.:58.00   1st Qu.:33.00    1st Qu.:0.3333   Class :character
 Median :60.00   Median :53.00    Median :0.8491   Mode  :character
 Mean   :56.64   Mean   :49.24    Mean   :1.1292
 3rd Qu.:68.00   3rd Qu.:68.00    3rd Qu.:1.6500
 Max.   :76.00   Max.   :76.00    Max.   :3.7586
 yield.plant_cull
 Min.   :0.00000
 1st Qu.:0.00000
 Median :0.01887
 Mean   :0.05404
 3rd Qu.:0.09091
 Max.   :0.36207
```

```
head(yield_df)
```

```
# A tibble: 6 × 17
   year site    treatment        planting harvest num_harvest `damaged add`
  <dbl> <chr>   <chr>               <dbl>   <dbl>       <dbl>         <dbl>
1  2025 wooster not supplemented        1       1          61            61
2  2025 wooster supplemented            1       2          42            42
3  2025 wooster supplemented            1       1          54            54
4  2025 wooster not supplemented        1       2          11            13
5  2025 wooster supplemented            1       2           9             9
6  2025 wooster not supplemented        1       2           8             0
# ℹ 10 more variables: weight_harvest <chr>, Cull_poll <dbl>,
#   Cull_poll_weight <chr>, Cull_other <dbl>, Cull_other_weight <chr>,
#   stand <dbl>, `functional stand` <dbl>, yield.plant <dbl>,
#   yield.plant_weight <chr>, yield.plant_cull <dbl>
```

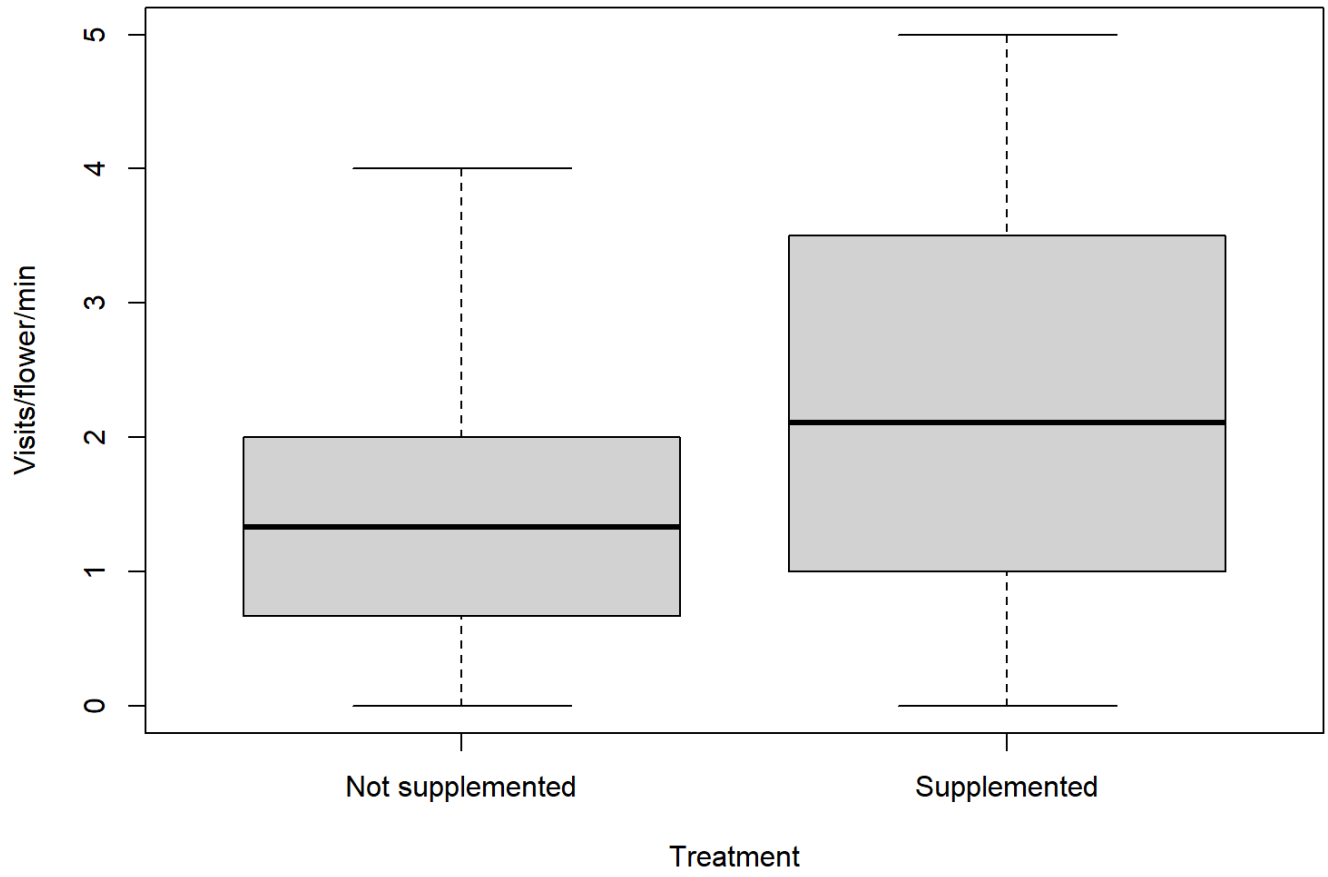
```
yield_df$treatment <- as.factor(yield_df$treatment)
str(yield_df)
```

```
tibble [25 × 17] (S3: tbl_df/tbl/data.frame)
 $ year              : num [1:25] 2025 2025 2025 2025 2025 ...
 $ site              : chr [1:25] "wooster" "wooster" "wooster" "wooster" ...
 $ treatment         : Factor w/ 2 levels "not supplemented",..: 1 2 2 1 2 1 2 1 2 1 ...
 $ planting          : num [1:25] 1 1 1 1 1 1 1 1 1 1 ...
 $ harvest           : num [1:25] 1 2 1 2 2 2 3 3 3 3 ...
```

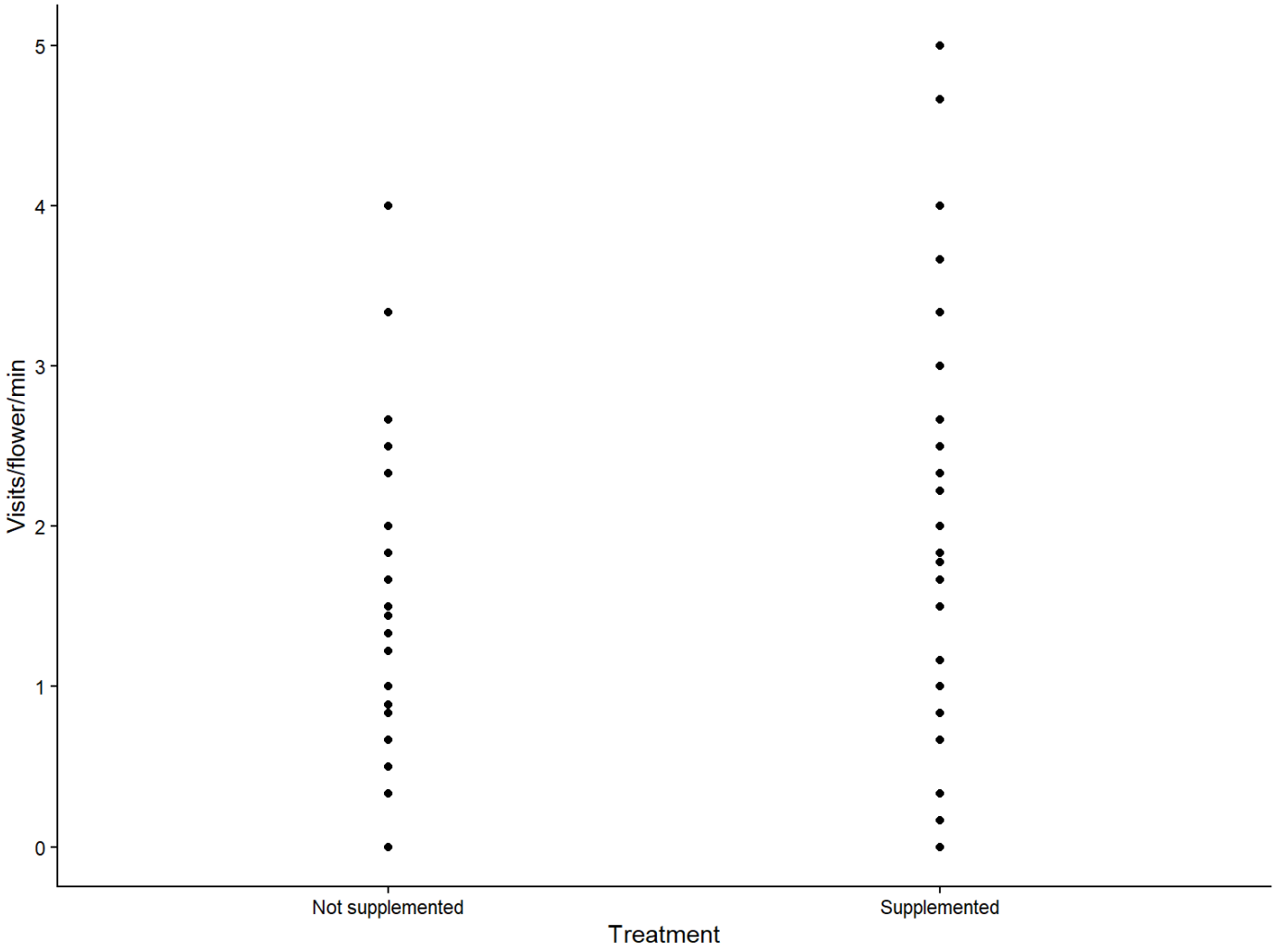
```
 $ num_harvest       : num [1:25] 61 42 54 11 9 8 9 7 15 33 ...
 $ damaged add       : num [1:25] 61 42 54 13 9 0 0 0 0 0 ...
 $ weight_harvest    : chr [1:25] "53.329000000000001" "42.616" "29.36" "8.0828000000000007" ...
 $ Cull_poll         : num [1:25] 8 2 0 3 0 0 0 3 3 2 ...
 $ Cull_poll_weight  : chr [1:25] "1.6990000000000001" "0.34200000000000003" "0"
"0.55000000000000004" ...
 $ Cull_other        : num [1:25] 2 0 0 0 0 0 0 0 0 0 ...
 $ Cull_other_weight : chr [1:25] "0.28000000000000003" "0" "0" "0" ...
 $ stand             : num [1:25] 60 34 32 60 32 60 32 60 32 60 ...
 $ functional stand  : num [1:25] 33 34 32 33 32 33 32 33 32 20 ...
 $ yield.plant       : num [1:25] 1.848 1.235 1.688 0.333 0.281 ...
 $ yield.plant_weight: chr [1:25] "1.6160303030000001" "1.2534117650000001" "0.91749999999999998"
"0.244933333" ...
 $ yield.plant_cull  : num [1:25] 0.2424 0.0588 0 0.0909 0 ...
```

Plot of visits/flower/min as a function of treatment:

```
plot(visits.flower.min~treatment, data= squash_bees_df, xlab= "Treatment",
         ylab="Visits/flower/min")
```

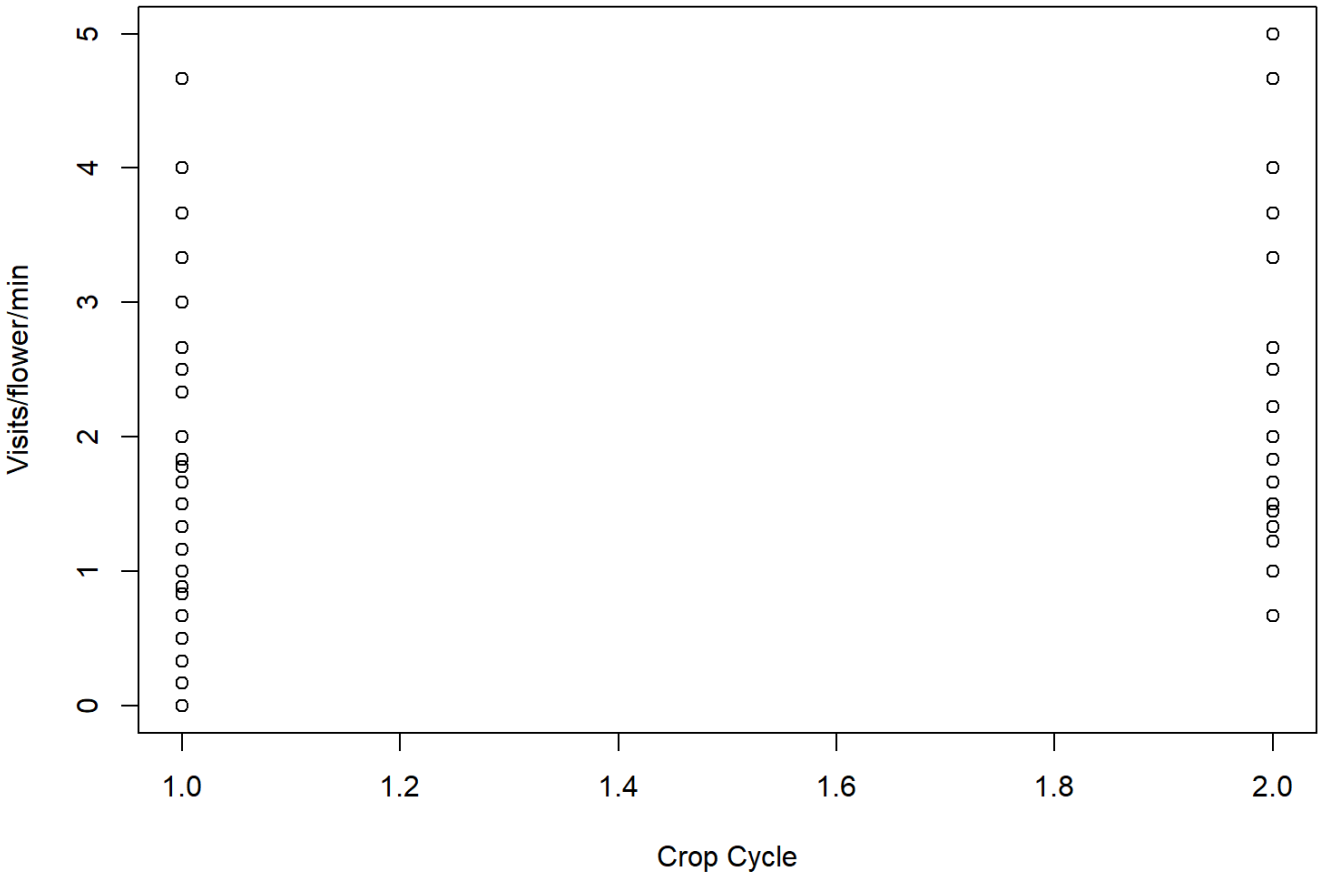


```
ggplot(data=squash_bees_df, mapping=aes(x=treatment , y=visits.flower.min)) +
geom_point()+theme_classic()+
  xlab("Treatment")+
  ylab("Visits/flower/min")
```

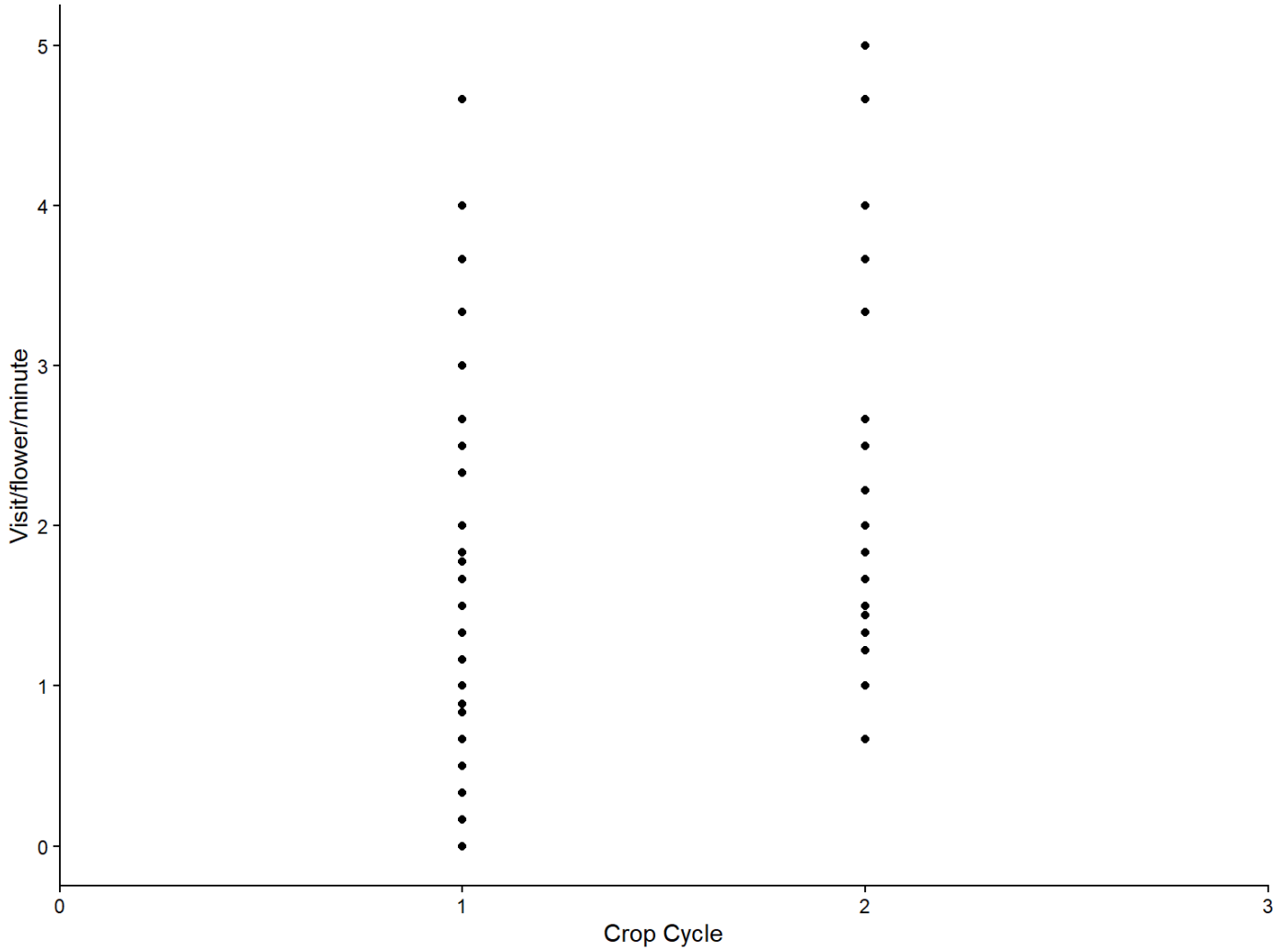


Plot of visits/flower/min as a function of planting date (~proxy for bloom date):

```
plot(visits.flower.min~planting, data= squash_bees_df, xlab="Crop Cycle", ylab=
        "Visits/flower/min")
```

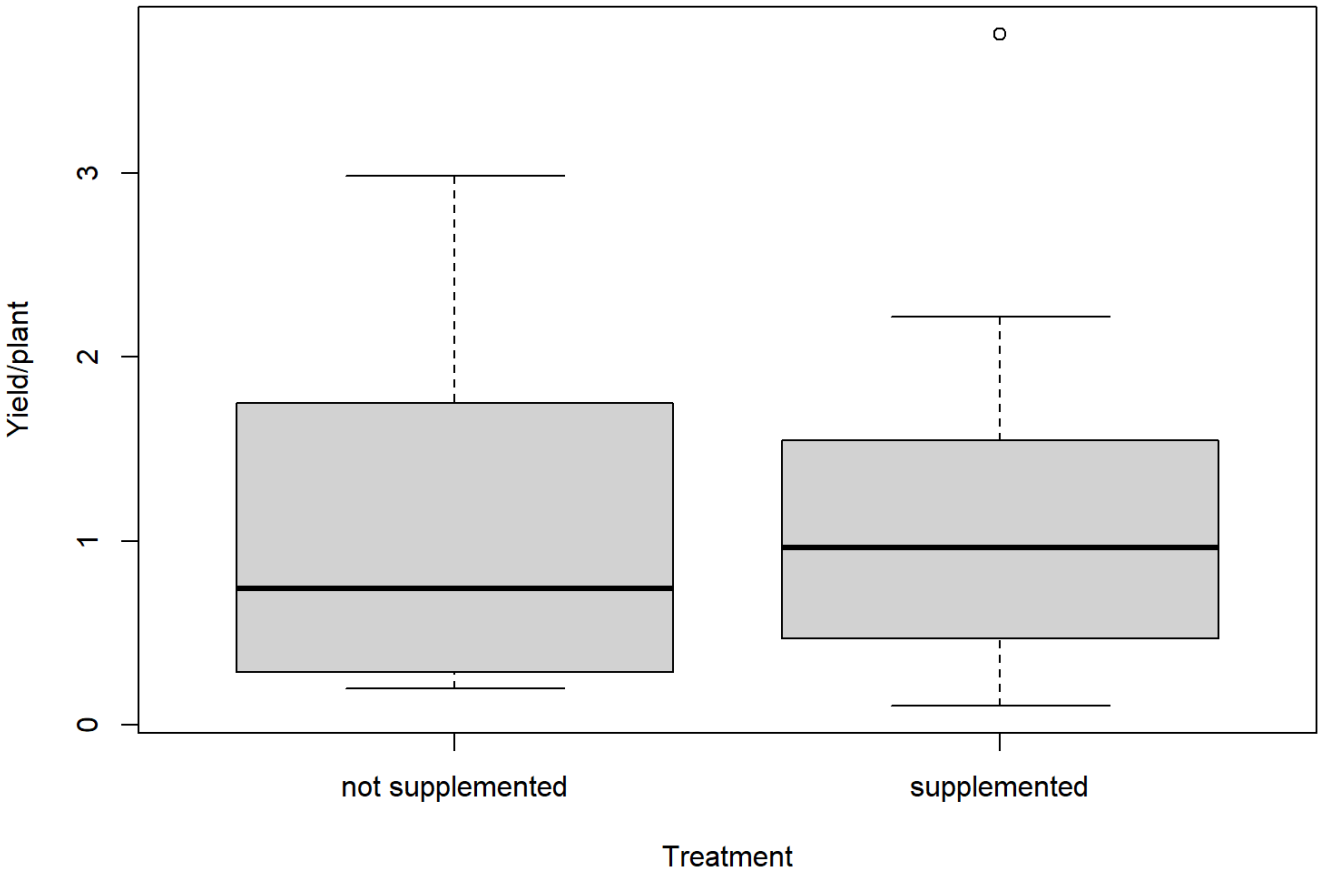



```
ggplot(data=squash_bees_df, mapping=aes(x=planting, y=visits.flower.min)) +
  geom_point()+theme_classic()+
  scale_y_continuous(limits = c(0, 5), breaks = seq(0, 5, by = 1)) +
  ylab("Visit/flower/minute")+
  scale_x_continuous(limits = c(0, 3), breaks = seq(0, 3, by = 1),
  expand=c(0,0))+
  xlab("Crop Cycle")
```

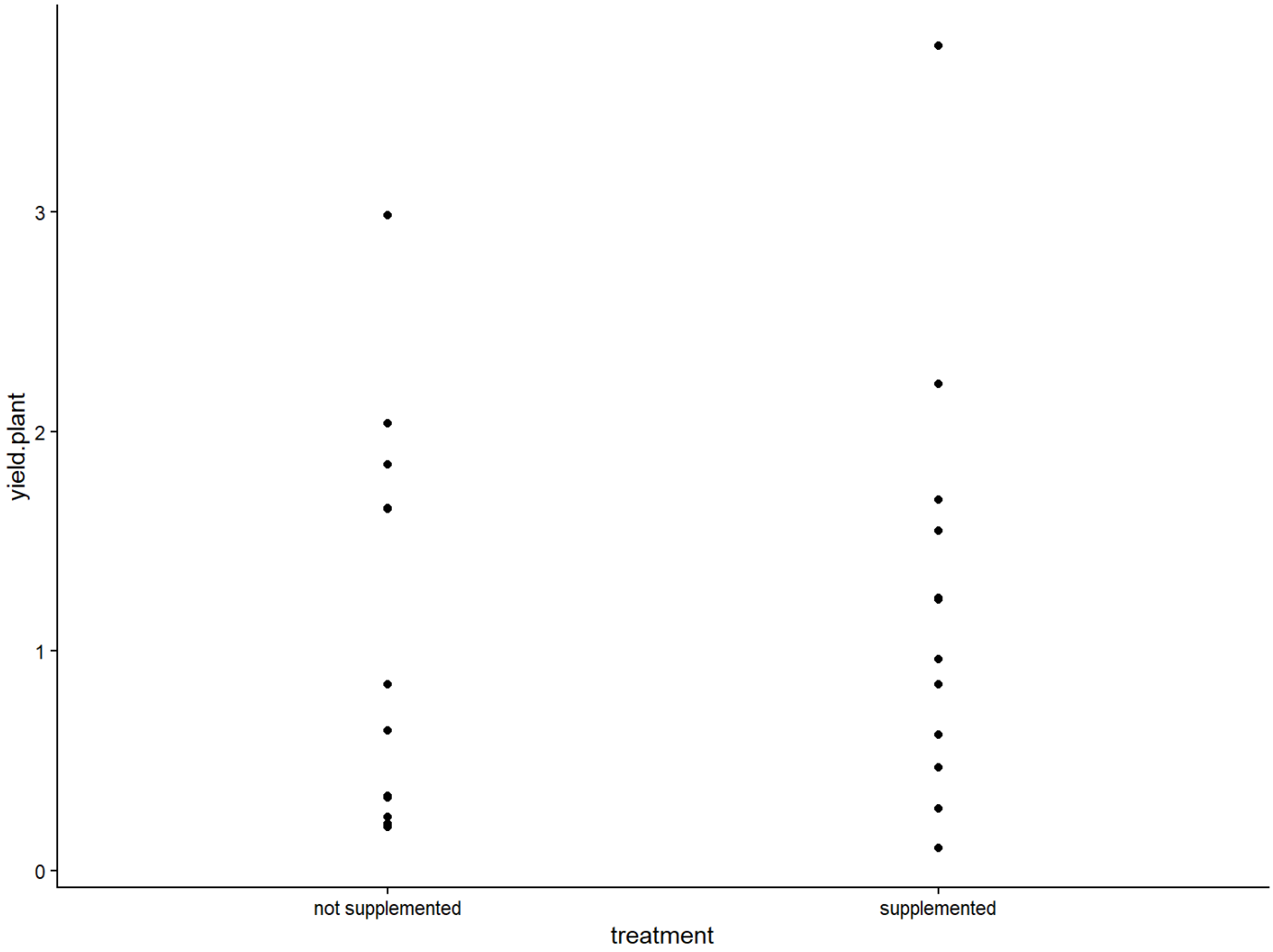


Plot of yield/plant as a function of treatment:

```
plot(yield.plant~treatment, data= yield_df, xlab= "Treatment", ylab= "Yield/plant")
```

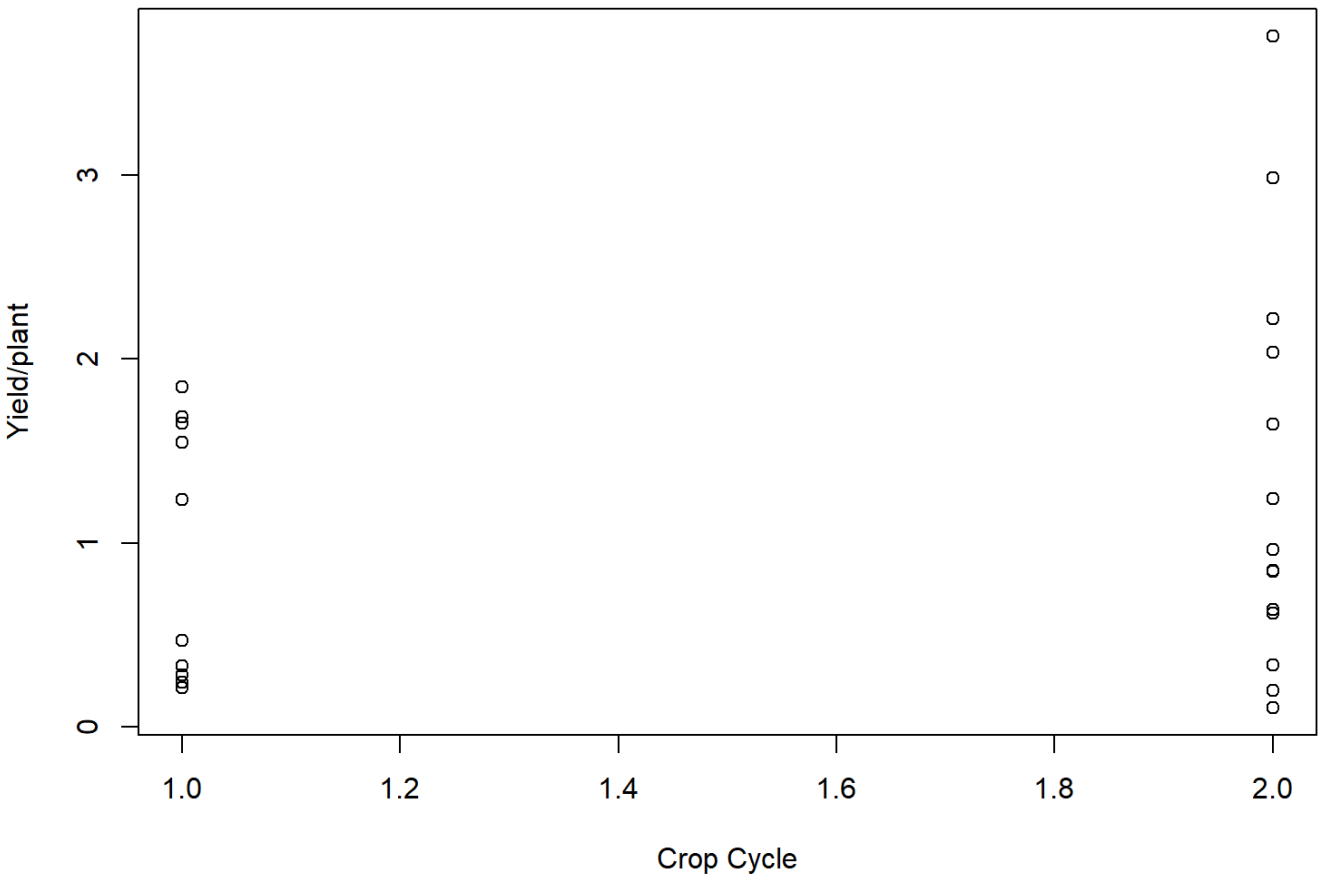



```
ggplot(data=yield_df, mapping=aes(x=treatment, y=yield.plant)) +
  geom_point()+theme_classic()
```

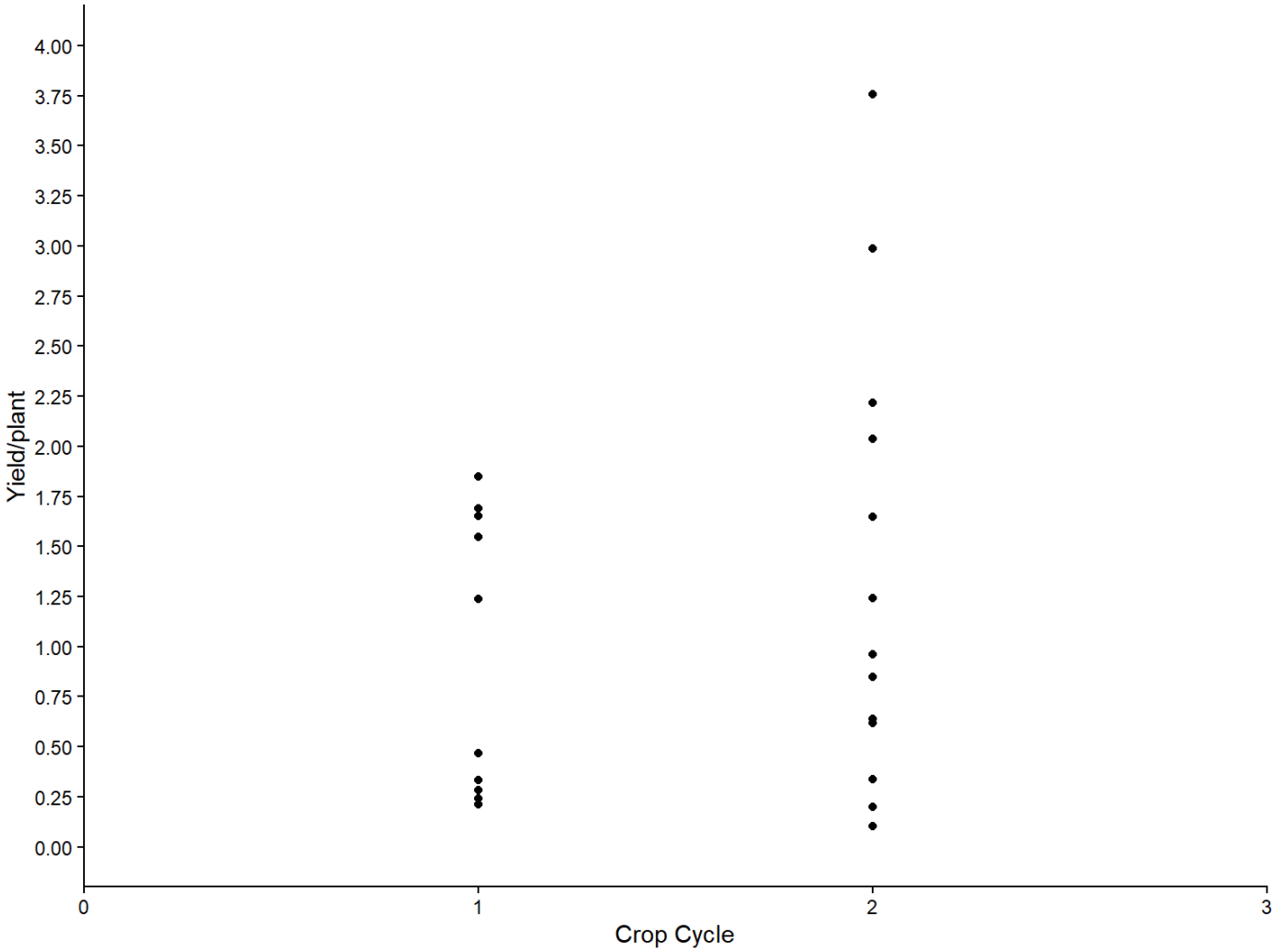


Plot of yield/plant as a function of planting date (~proxy for bloom date):

```
plot(yield.plant~planting, data= yield_df, xlab= "Crop Cycle", ylab= "Yield/plant")
```




```
ggplot(data=yield_df, mapping=aes(x=planting, y=yield.plant)) +
  geom_point()+ theme_classic()+
  scale_y_continuous(limits = c(0, 4), breaks = seq(0, 4, by = 0.25)) +
  ylab("Yield/plant")+
  scale_x_continuous(limits = c(0, 3), breaks = seq(0, 3, by = 1),
  expand=c(0,0))+
  xlab("Crop Cycle")
```


Yield/plant
4.00
3.75
3.50
3.25
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0
1
2
3
Crop Cycle