

BIO201 final presentation

Kasey Neff, Yilin Guan, Peter Nam

12.9.19

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(tidyverse)
library(readxl)
library(broom)
library(cowplot)
library(agricolae)
library(phyloseq)
library(tinytex)
set.seed(7)
```

Reserach Question: What supplements cause increase in propionate? Is this increase (or change) in propionate correlated to other changes in SCFA or variations in species of bacteria?

```
scfa_weekly <- read_delim(file = "~/Documents/Bio201_Project/final_project_data/DB_v_0.08/SCFA_wkly.txt",
                          delim = "\t", col_names = TRUE, trim_ws = TRUE,
                          na = c("", "NA")) %>%
  rename_all(tolower) %>%
  filter(quantity_compliant == "yes" | quantity_compliant == "none", study_week == "week1" | study_week == "week3") %>%
  select(-starts_with("ace"), -starts_with("but"), -ends_with("median")) %>%
  spread(study_week, propionate_mean) # separate propionate measurements from week 1 and week 3
```

```
## Parsed with column specification:
## cols(
##   Participant_ID = col_character(),
##   Study_week = col_character(),
##   Frequency = col_character(),
##   Semester = col_character(),
##   Supplement_consumed = col_character(),
##   Quantity_compliant = col_character(),
##   Acetate_median = col_double(),
##   Acetate_mean = col_double(),
##   Butyrate_median = col_double(),
##   Butyrate_mean = col_double(),
##   Propionate_median = col_double(),
##   Propionate_mean = col_double()
## )
```

```

# Filter for individuals who consumed Accessible
Accessible_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "Accessible")

# Filter for individuals who consumed no supplement
none_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "none")

# Filter for individuals who consumed BRMPS
BRMPS_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "BRMPS")

# Filter for individuals who consumed Arabino
Arabino_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "Arabino")

# Filter for individuals who consumed HiMaize
HiMaize_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "HiMaize")

# Filter for individuals who consumed Inulin
Inulin_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "Inulin")

# Filter for individuals who consumed Psyllium and BRMPS
Psyllium_BRMPS_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "Psyllium+BRMPS")

# Filter for individuals who consumed Psyllium
Psyllium_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "Psyllium")

# Filter for individuals who consumed BRMPS and Accessible
BRMPS_Accessible_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "BRMPS+Accessible")

# Filter for individuals who consumed LOODAT
LOODAT_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "LOODAT")

# Filter for individuals who consumed transition HiMaize
transition_HiMaize_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%

```

```

    filter (supplement_consumed == "transition_HiMaize")

# Filter for individuals who consumed HiMaize and BRMPS
HiMaize_BRMPS_weekly <- scfa_weekly %>%
  filter(!is.na(supplement_consumed)) %>%
  filter (supplement_consumed == "HiMaize+BRMPS")

# Saved data frames
write_delim(Accessible_weekly, path = "~/Documents/Bio201_Project/curated_data/Accessible_weekly.txt", delim = "\t")
write_delim(none_weekly, path = "~/Documents/Bio201_Project/curated_data/none_weekly.txt", delim = "\t")
write_delim(BRMPS_weekly, path = "~/Documents/Bio201_Project/curated_data/BRMPS_weekly.txt", delim = "\t")
write_delim(HiMaize_weekly, path = "~/Documents/Bio201_Project/curated_data/HiMaize_weekly.txt", delim = "\t")
write_delim(Inulin_weekly, path = "~/Documents/Bio201_Project/curated_data/Inulin_weekly.txt", delim = "\t")
write_delim(Psyllium_BRMPS_weekly, path = "~/Documents/Bio201_Project/curated_data/Psyllium_BRMPS_weekly.txt", delim = "\t")
write_delim(Psyllium_weekly, path = "~/Documents/Bio201_Project/curated_data/Psyllium_weekly.txt", delim = "\t")
write_delim(BRMPS_Accessible_weekly, path = "~/Documents/Bio201_Project/curated_data/BRMPS_Accessible_weekly.txt", delim = "\t")
write_delim(LOODAT_weekly, path = "~/Documents/Bio201_Project/curated_data/LOODAT_weekly.txt", delim = "\t")
write_delim(transition_HiMaize_weekly, path = "~/Documents/Bio201_Project/curated_data/transition_HiMaize_weekly.txt", delim = "\t")
write_delim(HiMaize_BRMPS_weekly, path = "~/Documents/Bio201_Project/curated_data/HiMaize_BRMPS_weekly.txt", delim = "\t")

```

Assumptions and T-Test for Supplements

```

# Accessible
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "Accessible") %>%
  summarize(sample_size = n()) # week1 n=31 & week3 n=32, not use the NA

```

```

## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1           32

```

```

# Normality
shapiro.test(Accessible_weekly$week1)

```

```

##
##  Shapiro-Wilk normality test
##
## data:  Accessible_weekly$week1
## W = 0.86726, p-value = 0.001208

```

```

ggplot(Accessible_weekly, aes(x = week1)) + geom_histogram() # p-value = 0.001208, skewed to the right

```

```

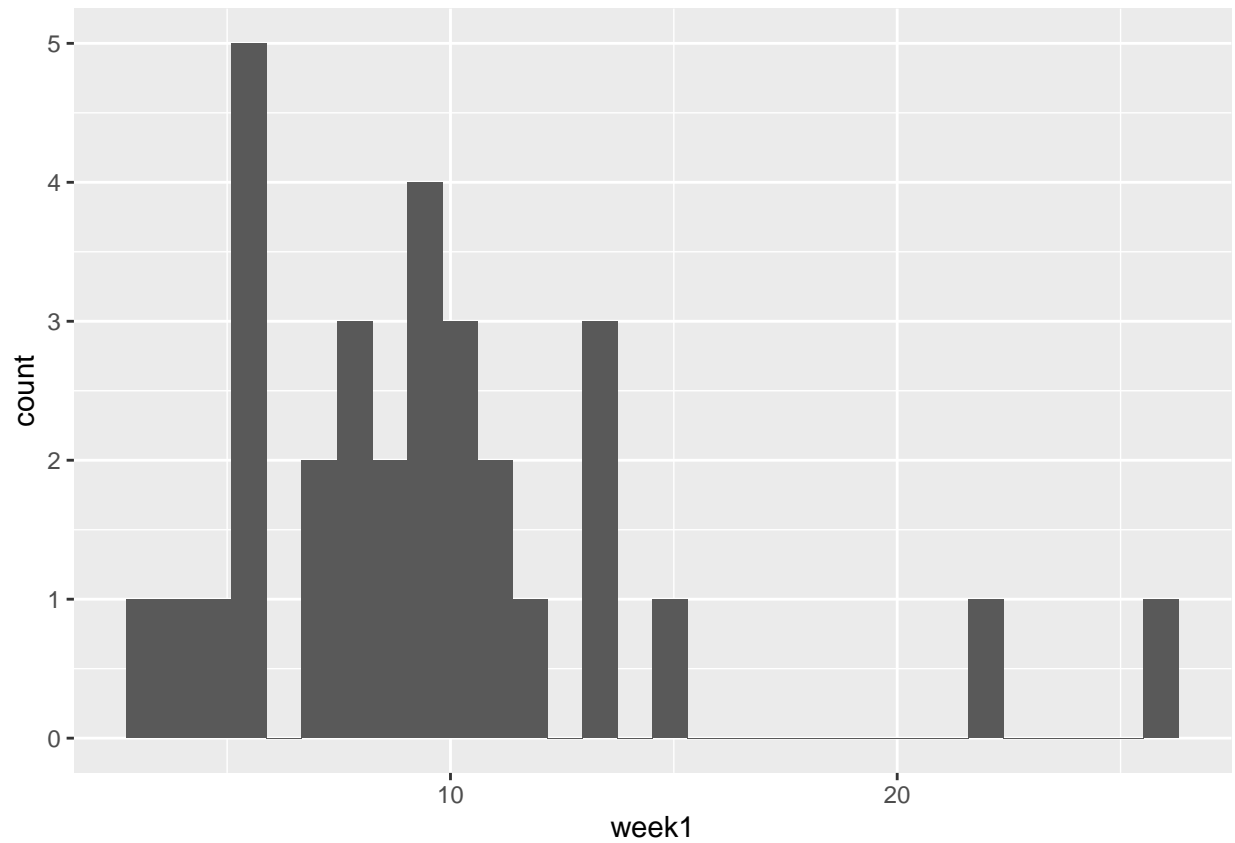
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

```

```

## Warning: Removed 1 rows containing non-finite values (stat_bin).

```

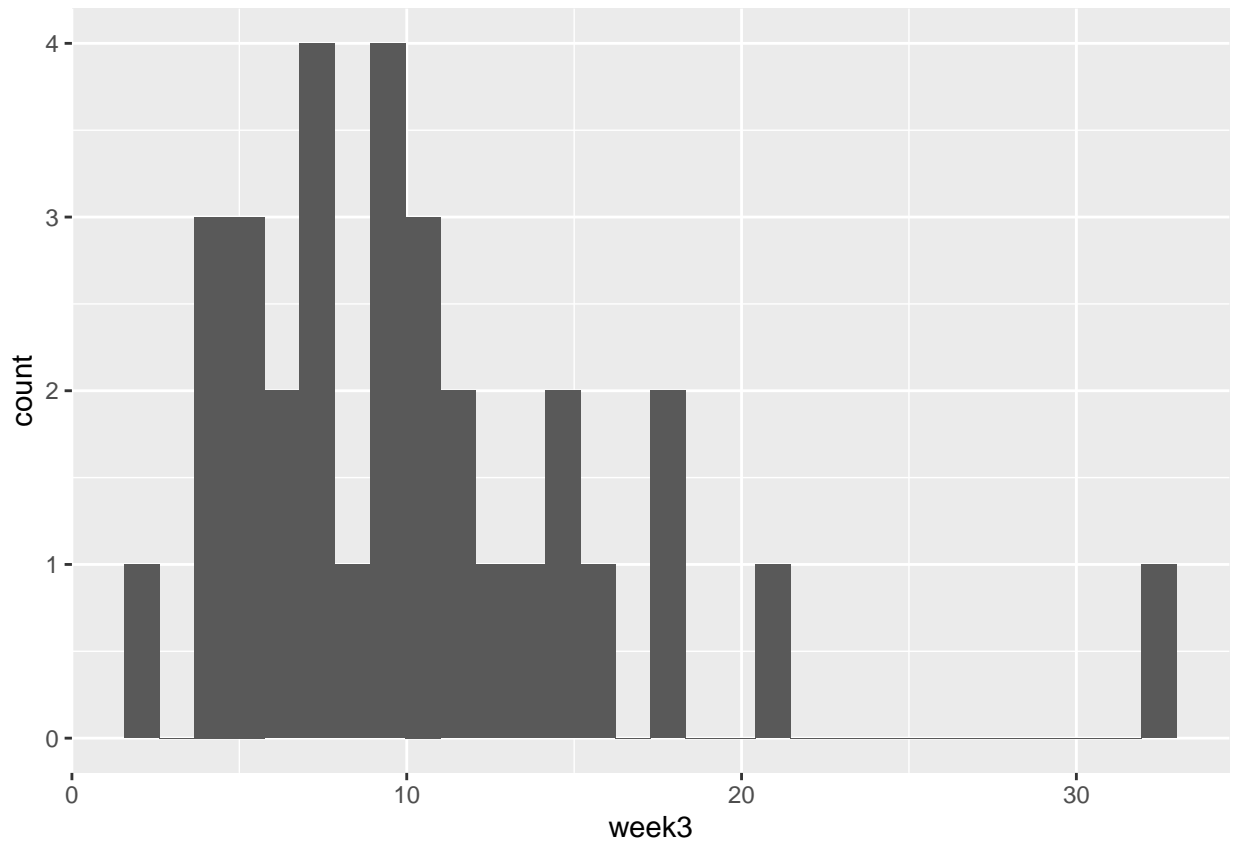


```
shapiro.test(Accessible_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Accessible_weekly$week3
## W = 0.85986, p-value = 0.0006834
```

```
ggplot(Accessible_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.0006834, skewed to the right
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
# Equal Variance
var.test(x = Accessible_weekly$week1,
         y = Accessible_weekly$week3,
         alternative = "two.sided") # p-value = 0.2249, variances are not equal
```

```
##
## F test to compare two variances
##
## data: Accessible_weekly$week1 and Accessible_weekly$week3
## F = 0.64004, num df = 30, denom df = 31, p-value = 0.2249
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.3111788 1.3221543
## sample estimates:
## ratio of variances
## 0.6400405
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
Accessible_weekly <- na.omit(Accessible_weekly)
wilcox.test(x = Accessible_weekly$week1,
            y = Accessible_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE)
```

```
## Warning in wilcox.test.default(x = Accessible_weekly$week1, y =
## Accessible_weekly$week3, : cannot compute exact p-value with ties
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: Accessible_weekly$week1 and Accessible_weekly$week3
## V = 189, p-value = 0.1258
## alternative hypothesis: true location shift is less than 0
```

#the p value for the wilcox is 0.45, which is not less than 0.05 and fails to reject the null. There is

```
# Arabino
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "Arabino") %>%
  summarize(sample_size = n()) # week1 n=17 & week3 n=13, not use the NA -> WILCOXON
```

```
## # A tibble: 1 x 1
##   sample_size
##       <int>
## 1         21
```

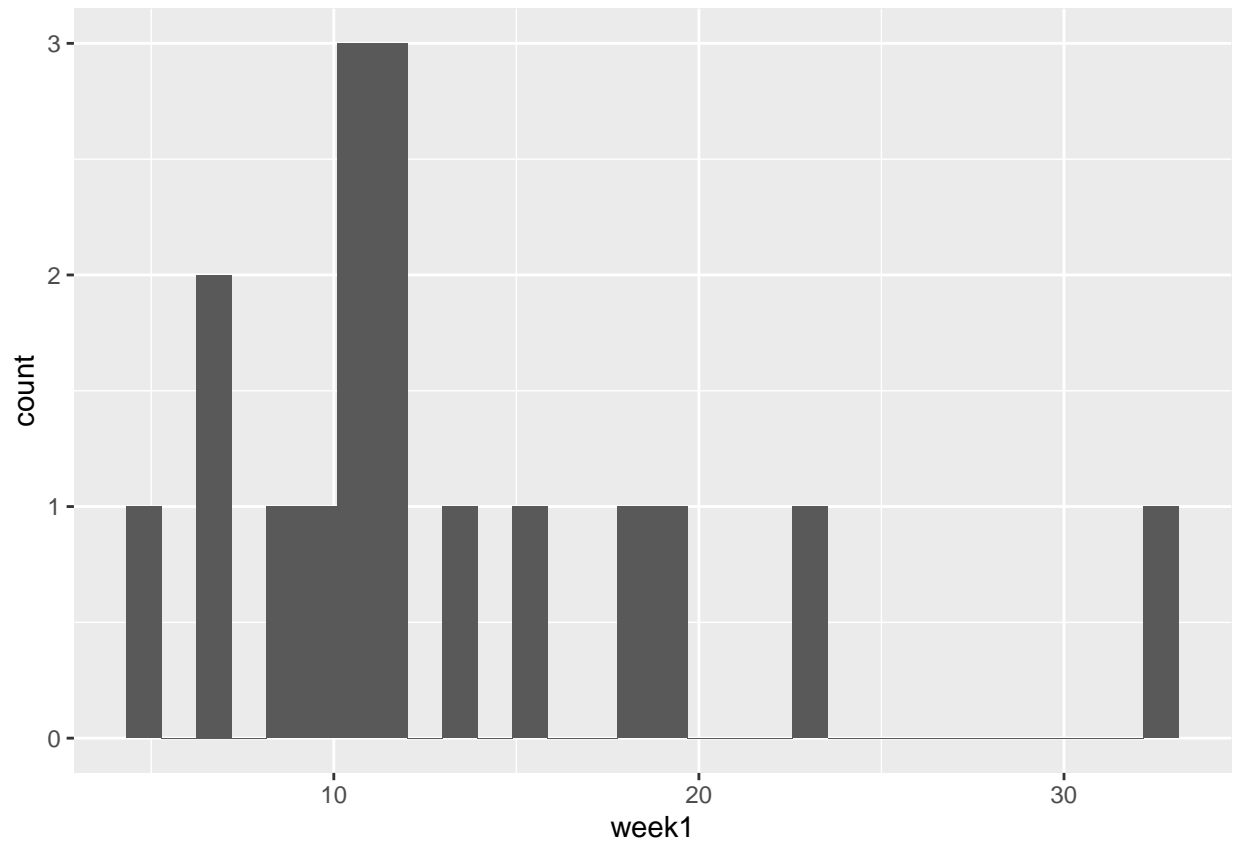
```
# Normality
shapiro.test(Arabino_weekly$week1)
```

```
##
## Shapiro-Wilk normality test
##
## data: Arabino_weekly$week1
## W = 0.85604, p-value = 0.01328
```

ggplot(Arabino_weekly, aes(x = week1)) + geom_histogram() # p-value = 0.01328, skewed to the right -> n

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 4 rows containing non-finite values (stat_bin).
```



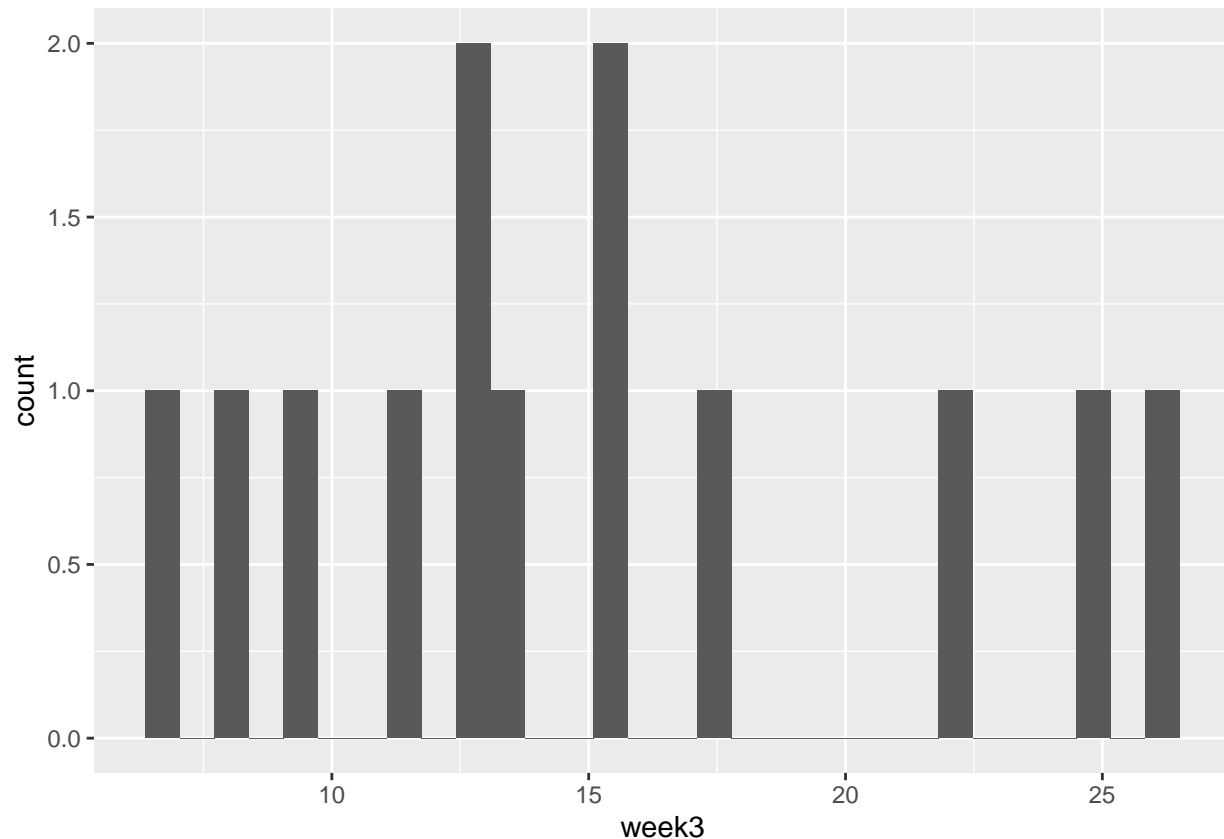
```
shapiro.test(Arabino_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Arabino_weekly$week3
## W = 0.94425, p-value = 0.5143
```

```
ggplot(Arabino_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.5143, sort of normal shaped ->
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 8 rows containing non-finite values (stat_bin).
```



```
# Equal Variance
var.test(x = Arabino_weekly$week1,
         y = Arabino_weekly$week3,
         alternative = "two.sided") # p-value = 0.6808, variances are equal
```

```
##
## F test to compare two variances
##
## data: Arabino_weekly$week1 and Arabino_weekly$week3
## F = 1.2739, num df = 16, denom df = 12, p-value = 0.6808
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.4042281 3.6804611
## sample estimates:
## ratio of variances
## 1.273936
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
Arabino_weekly <- na.omit(Arabino_weekly)
wilcox.test(x = Arabino_weekly$week1,
            y = Arabino_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE)
```

```
##
## Wilcoxon signed rank test
```



```
##
## data: Arabino_weekly$week1 and Arabino_weekly$week3
## V = 17, p-value = 0.0874
## alternative hypothesis: true location shift is less than 0

# p-value = 0.0874, which is larger than 0.05 we fail to reject the null. There is not sufficient evidence
# BRMPS
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "BRMPS") %>%
  summarize(sample_size = n()) # week1 n=151 & week3 n=151, not use the NA

## # A tibble: 1 x 1
##   sample_size
##       <int>
## 1         156

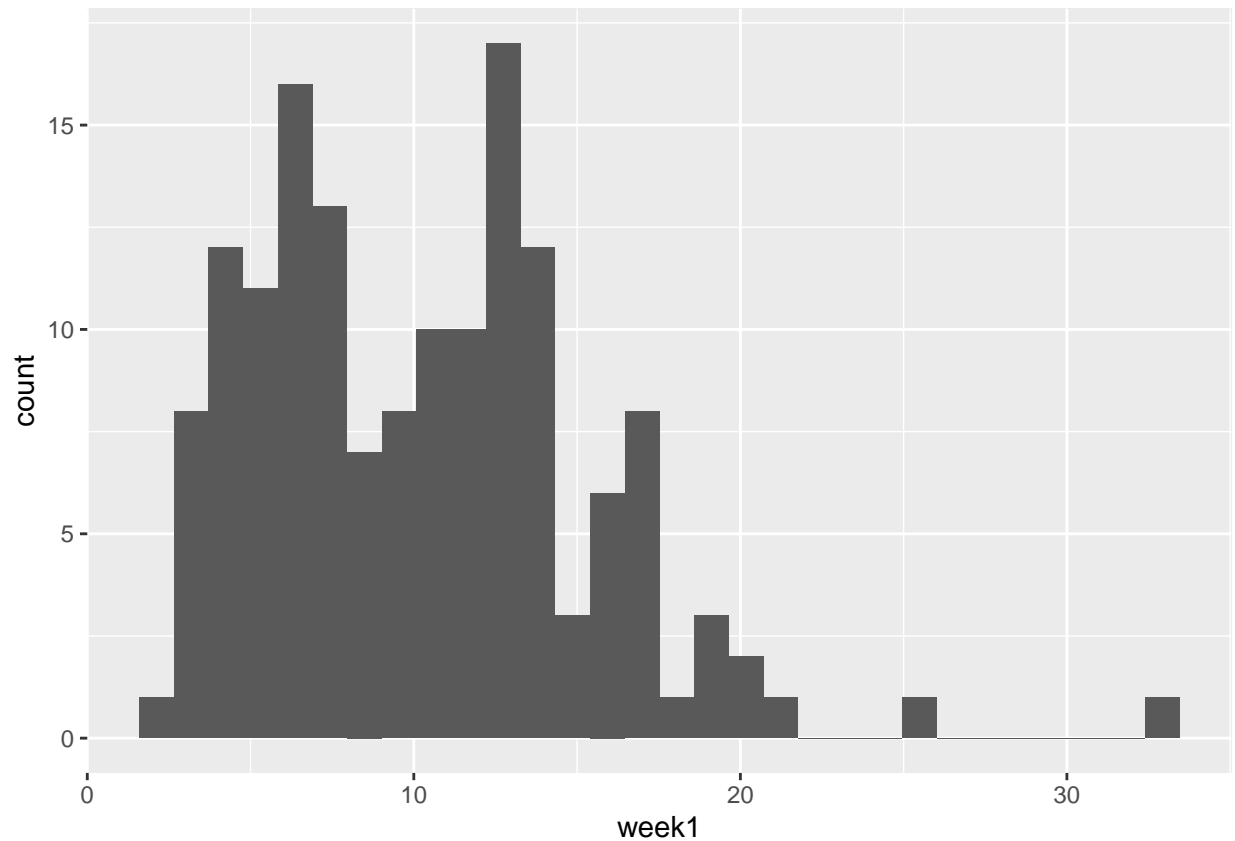
# Normality
shapiro.test(BRMPS_weekly$week1)

##
## Shapiro-Wilk normality test
##
## data: BRMPS_weekly$week1
## W = 0.94132, p-value = 6.225e-06

ggplot(BRMPS_weekly, aes(x = week1)) + geom_histogram() # p-value = 6.225e-06, bimodal and skewed to the right

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 5 rows containing non-finite values (stat_bin).
```



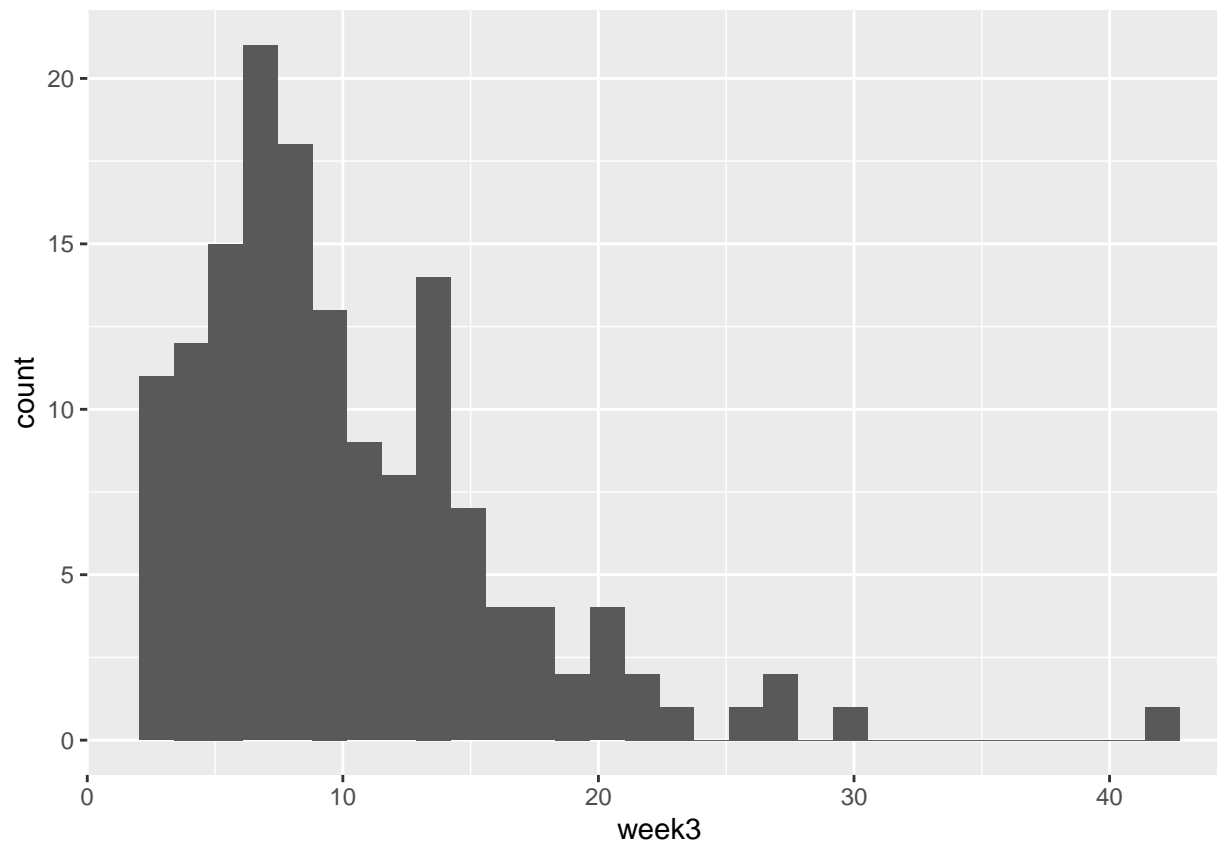
```
shapiro.test(BRMPS_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  BRMPS_weekly$week3
## W = 0.88151, p-value = 1.357e-09
```

```
ggplot(BRMPS_weekly, aes(x = week3)) + geom_histogram() # p-value = 1.357e-09, skewed to the right -> n
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 6 rows containing non-finite values (stat_bin).
```



```
# Equal Variance
var.test(x = BRMPS_weekly$week1,
         y = BRMPS_weekly$week3,
         alternative = "two.sided") # p-value = 0.01992, variances are not equal
```

```
##
## F test to compare two variances
##
## data: BRMPS_weekly$week1 and BRMPS_weekly$week3
## F = 0.68219, num df = 150, denom df = 149, p-value = 0.01992
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.4944289 0.9411002
## sample estimates:
## ratio of variances
## 0.6821941
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
BRMPS_weekly <- na.omit(BRMPS_weekly)
wilcox.test(x = BRMPS_weekly$week1,
            y = BRMPS_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE)
```

```
##
## Wilcoxon signed rank test with continuity correction
```

```
##
## data: BRMPS_weekly$week1 and BRMPS_weekly$week3
## V = 5793, p-value = 0.8386
## alternative hypothesis: true location shift is less than 0

# p-value = 0.8386, which is larger than 0.05 and we fail to reject the null. There is not sufficient evidence to reject the null.

# BRMPS + Accessible
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "BRMPS+Accessible") %>%
  summarize(sample_size = n()) # week1 n=17 & week3 n=16, not use the NA -> WILCOXON

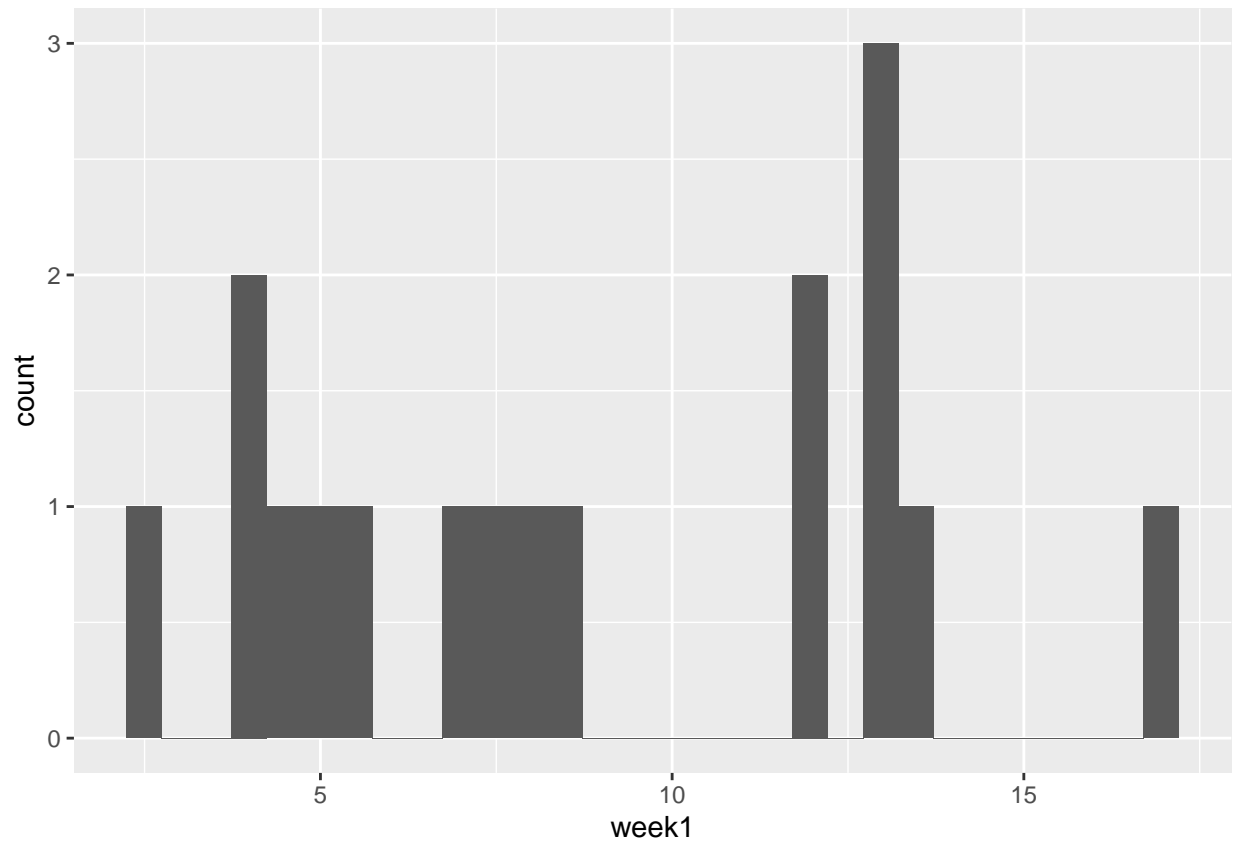
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1           17

# Normality
shapiro.test(BRMPS_Accessible_weekly$week1)

##
## Shapiro-Wilk normality test
##
## data: BRMPS_Accessible_weekly$week1
## W = 0.92806, p-value = 0.202

ggplot(BRMPS_Accessible_weekly, aes(x = week1)) + geom_histogram() # p-value = 0.202, sort of bimodal as expected

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



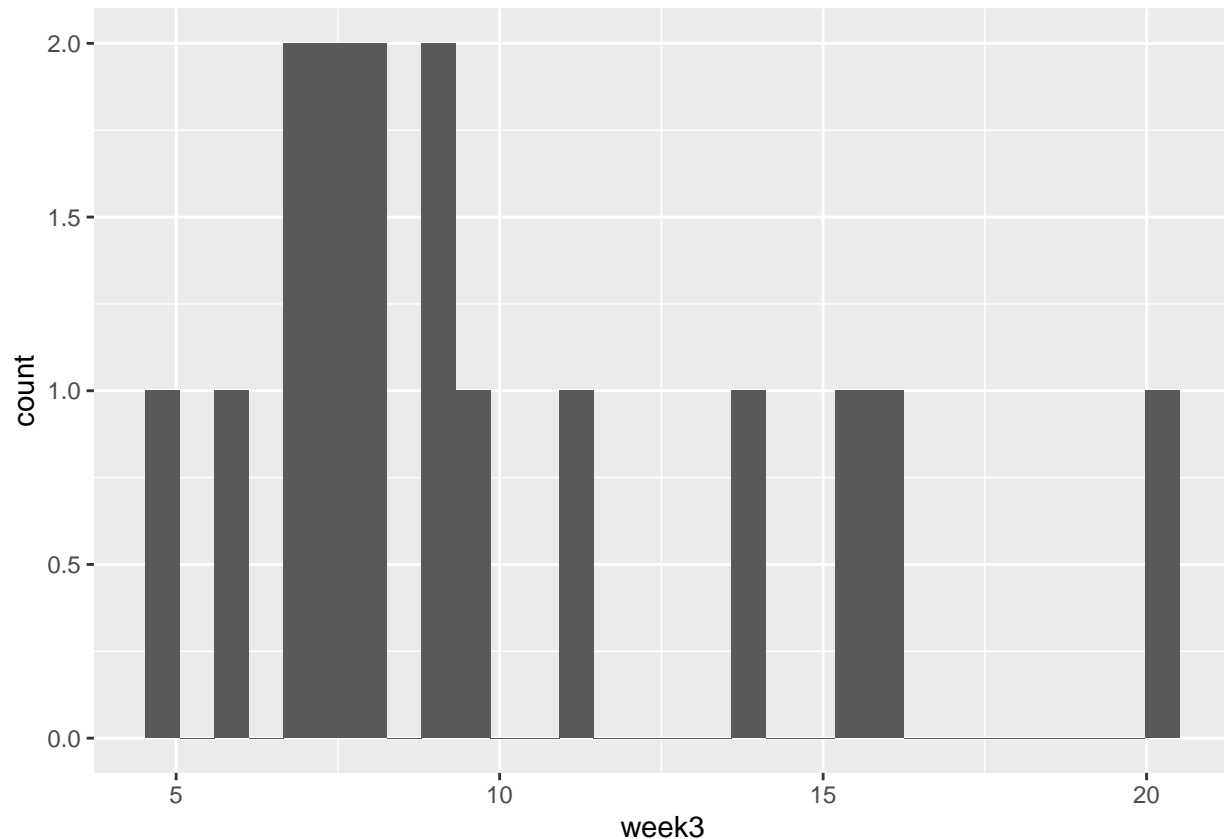
```
shapiro.test(BRMPS_Accessible_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  BRMPS_Accessible_weekly$week3
## W = 0.86964, p-value = 0.02685
```

```
ggplot(BRMPS_Accessible_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.02685, skewed to the r
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



```
# Equal Variance
var.test(x = BRMPS_Accessible_weekly$week1,
         y = BRMPS_Accessible_weekly$week3,
         alternative = "two.sided") # p-value = 0.9149, variances are equal
```

```
##
## F test to compare two variances
##
## data: BRMPS_Accessible_weekly$week1 and BRMPS_Accessible_weekly$week3
## F = 1.0598, num df = 16, denom df = 15, p-value = 0.9149
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.3736881 2.9542022
## sample estimates:
## ratio of variances
## 1.059797
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
BRMPS_Accessible_weekly <- na.omit(BRMPS_Accessible_weekly)
wilcox.test(x = BRMPS_Accessible_weekly$week1,
            y = BRMPS_Accessible_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE)
```

```
##
## Wilcoxon signed rank test
```

```
##
## data:  BRMPS_Accessible_weekly$week1 and BRMPS_Accessible_weekly$week3
## V = 65, p-value = 0.45
## alternative hypothesis: true location shift is less than 0

# p-value = 0.45, which is not less than 0.05 and we fail to reject the null. There is not sufficient evidence to reject the null hypothesis.

# HiMaize
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "HiMaize") %>%
  summarize(sample_size = n()) # week1 n=36 & week3 n=36, not use the NA

## # A tibble: 1 x 1
##   sample_size
##       <int>
## 1         43

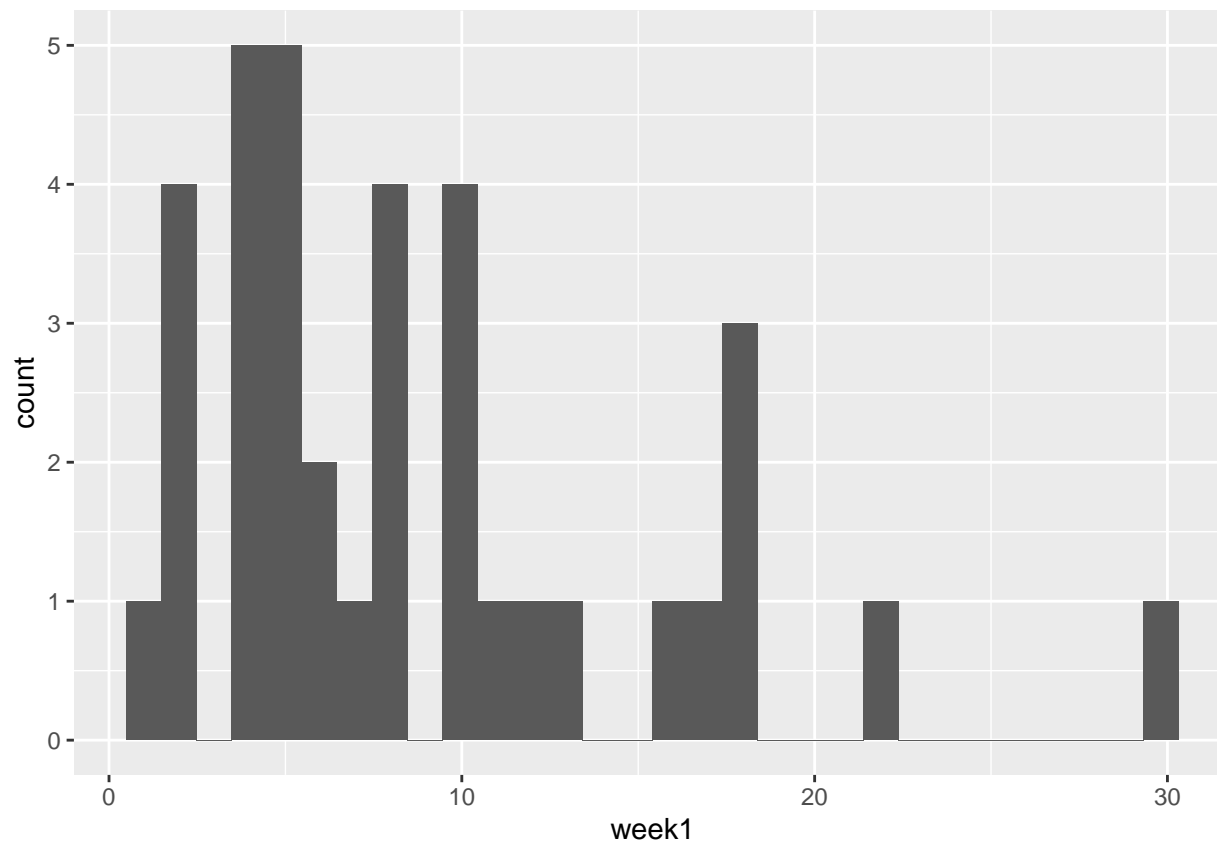
# Normality
shapiro.test(HiMaize_weekly$week1)

##
## Shapiro-Wilk normality test
##
## data:  HiMaize_weekly$week1
## W = 0.88161, p-value = 0.001113

ggplot(HiMaize_weekly, aes(x = week1)) + geom_histogram() # p-value = 0.001113, skewed to the right -> not normal

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 7 rows containing non-finite values (stat_bin).
```



```
shapiro.test(HiMaize_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  HiMaize_weekly$week3
## W = 0.87984, p-value = 0.001003
```

```
graph_Himaize <- ggplot(HiMaize_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.001003, skewed
# Equal Variance
var.test(x = HiMaize_weekly$week1,
         y = HiMaize_weekly$week3,
         alternative = "two.sided") # p-value = 0.1558, variances are equal
```

```
##
##  F test to compare two variances
##
## data:  HiMaize_weekly$week1 and HiMaize_weekly$week3
## F = 0.61535, num df = 35, denom df = 35, p-value = 0.1558
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3137789 1.2067535
## sample estimates:
## ratio of variances
##      0.6153485
```



```
# Statistical Test (Wilcoxon because data is nonparametric)
HiMaize_weekly <- na.omit(HiMaize_weekly)
wilcox.test(x = HiMaize_weekly$week1,
            y = HiMaize_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE)
```

```
## Warning in wilcox.test.default(x = HiMaize_weekly$week1, y =
## HiMaize_weekly$week3, : cannot compute exact p-value with ties
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: HiMaize_weekly$week1 and HiMaize_weekly$week3
## V = 122.5, p-value = 0.01215
## alternative hypothesis: true location shift is less than 0
```

```
# p-value = 0.01215, which is less than 0.05 and we reject the null. There is evidence that the propion
```

```
# HiMaize + BRMPS
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "HiMaize+BRMPS") %>%
  summarize(sample_size = n()) # week1 n=21 & week3 n=22, not use the NA - WILCOX
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1         22
```

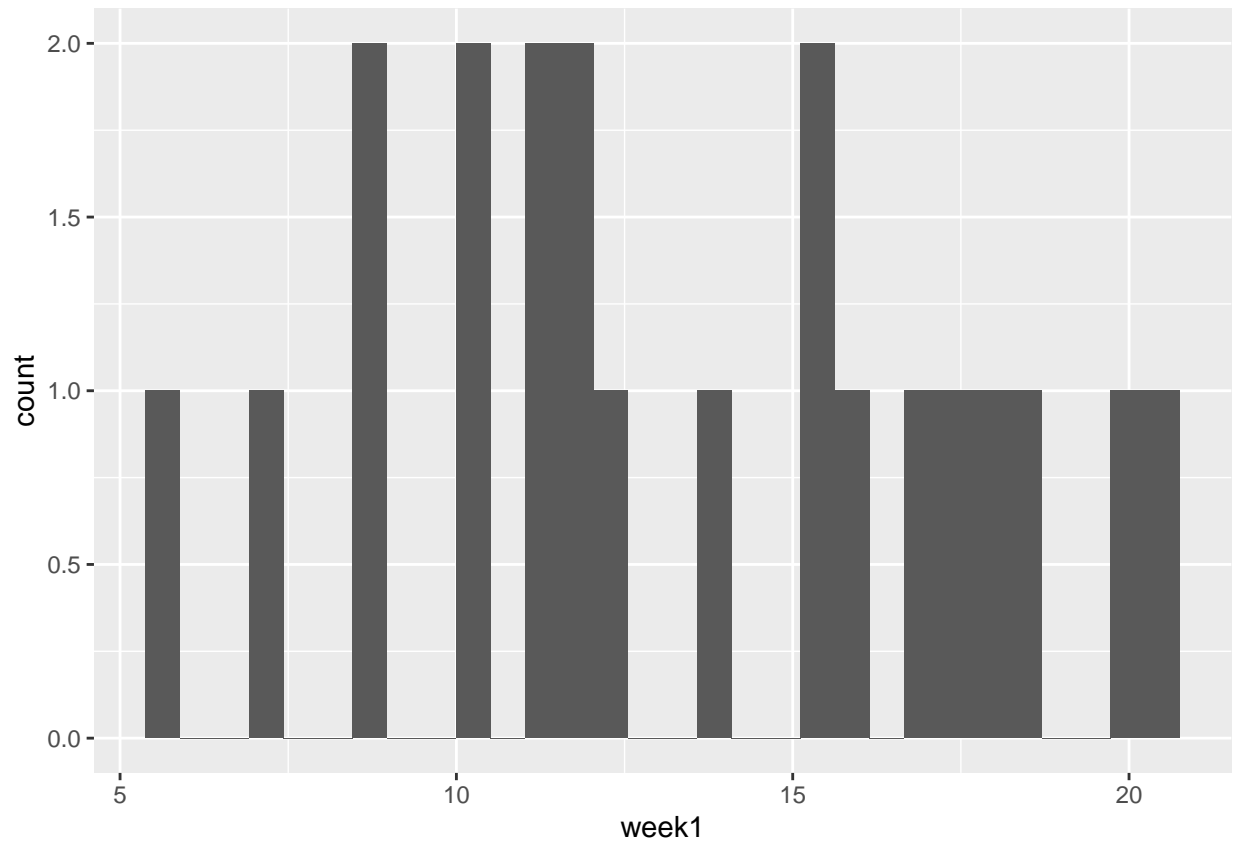
```
# Normality
shapiro.test(HiMaize_BRMPS_weekly$week1)
```

```
##
## Shapiro-Wilk normality test
##
## data: HiMaize_BRMPS_weekly$week1
## W = 0.96967, p-value = 0.7258
```

```
ggplot(HiMaize_BRMPS_weekly, aes(x = week1)) + geom_histogram() # p-value = 0.7258, normal shape -> yet
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```

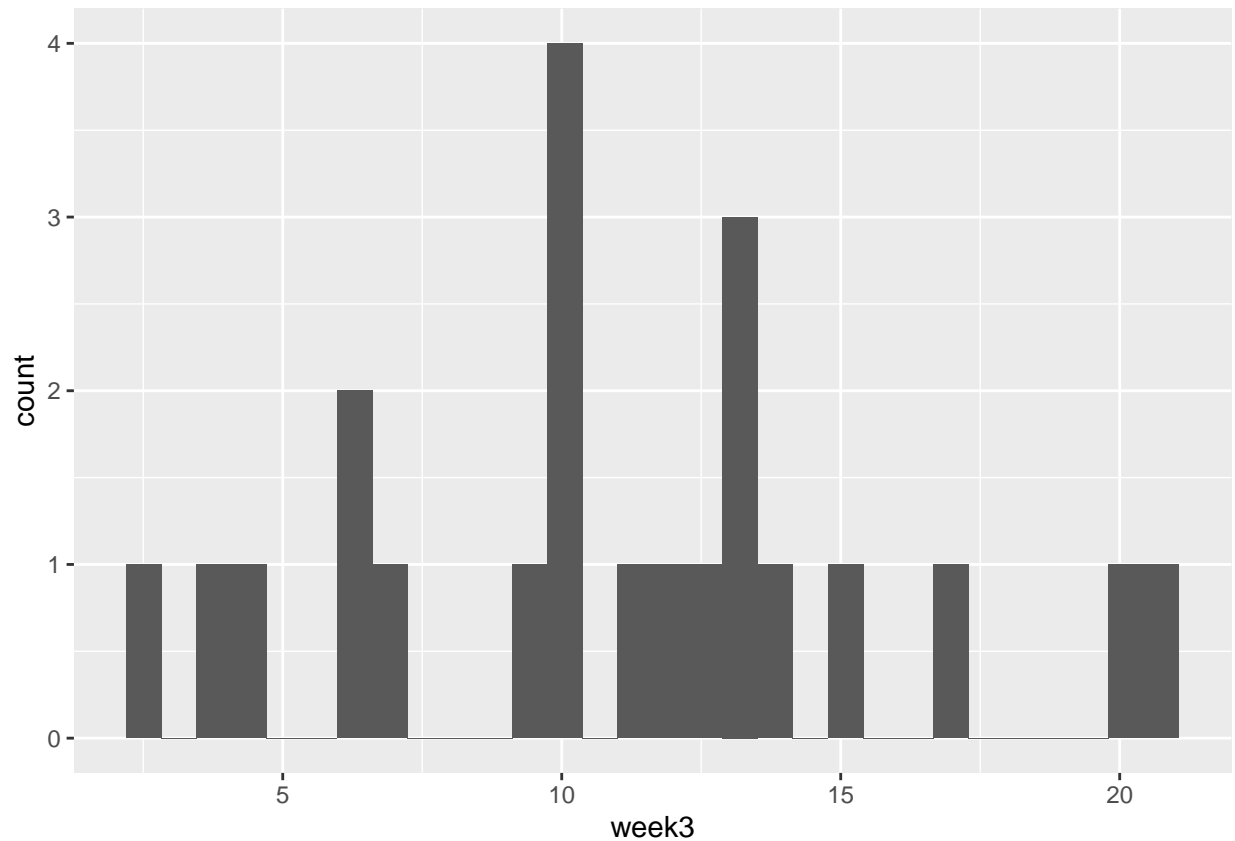


```
shapiro.test(HiMaize_BRMPS_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  HiMaize_BRMPS_weekly$week3
## W = 0.97454, p-value = 0.8131
```

```
ggplot(HiMaize_BRMPS_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.8131, normal shape -> yet
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
# Equal Variance
var.test(x = HiMaize_BRMPs_weekly$week1,
         y = HiMaize_BRMPs_weekly$week3,
         alternative = "two.sided") # p-value = 0.5405, variances are equal
```

```
##
## F test to compare two variances
##
## data: HiMaize_BRMPs_weekly$week1 and HiMaize_BRMPs_weekly$week3
## F = 0.75873, num df = 20, denom df = 21, p-value = 0.5405
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.312914 1.857694
## sample estimates:
## ratio of variances
## 0.7587337
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
HiMaize_BRMPs_weekly <- na.omit(HiMaize_BRMPs_weekly)
wilcox.test(x = HiMaize_BRMPs_weekly$week1,
            y = HiMaize_BRMPs_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE)
```

```
##
## Wilcoxon signed rank test
```

```
##
## data: HiMaize_BRMPS_weekly$week1 and HiMaize_BRMPS_weekly$week3
## V = 185, p-value = 0.9937
## alternative hypothesis: true location shift is less than 0

# p-value = 0.9937, which is larger than 0.05 and we fail to reject the null. There is not sufficient evidence to reject the null hypothesis.

# Inulin
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "Inulin") %>%
  summarize(sample_size = n()) # week1 n=44 & week3 n=41, not use the NA

## # A tibble: 1 x 1
##   sample_size
##       <int>
## 1         48

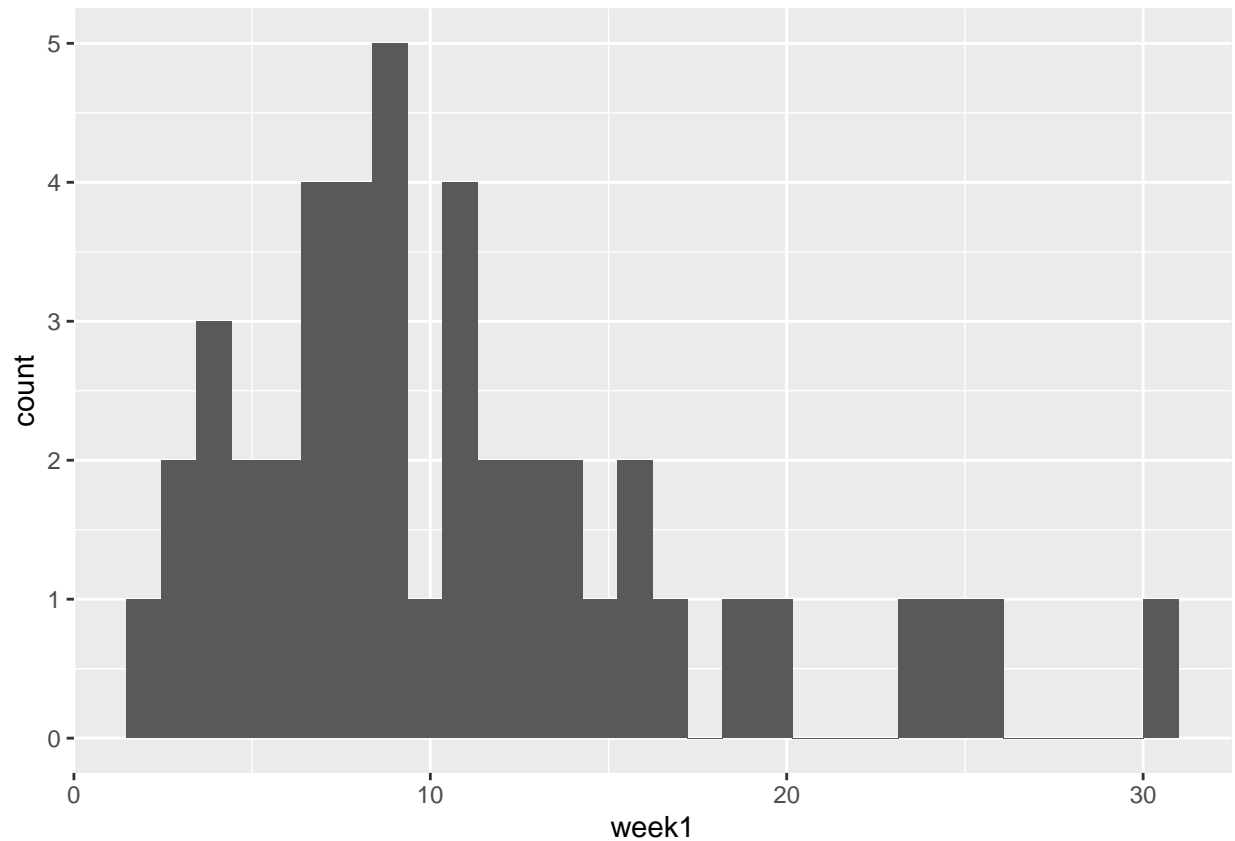
# Normality
shapiro.test(Inulin_weekly$week1)

##
## Shapiro-Wilk normality test
##
## data: Inulin_weekly$week1
## W = 0.90826, p-value = 0.001972

ggplot(Inulin_weekly, aes(x = week1)) + geom_histogram() # p-value = 0.001972, skewed to the right -> not normal

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 4 rows containing non-finite values (stat_bin).
```



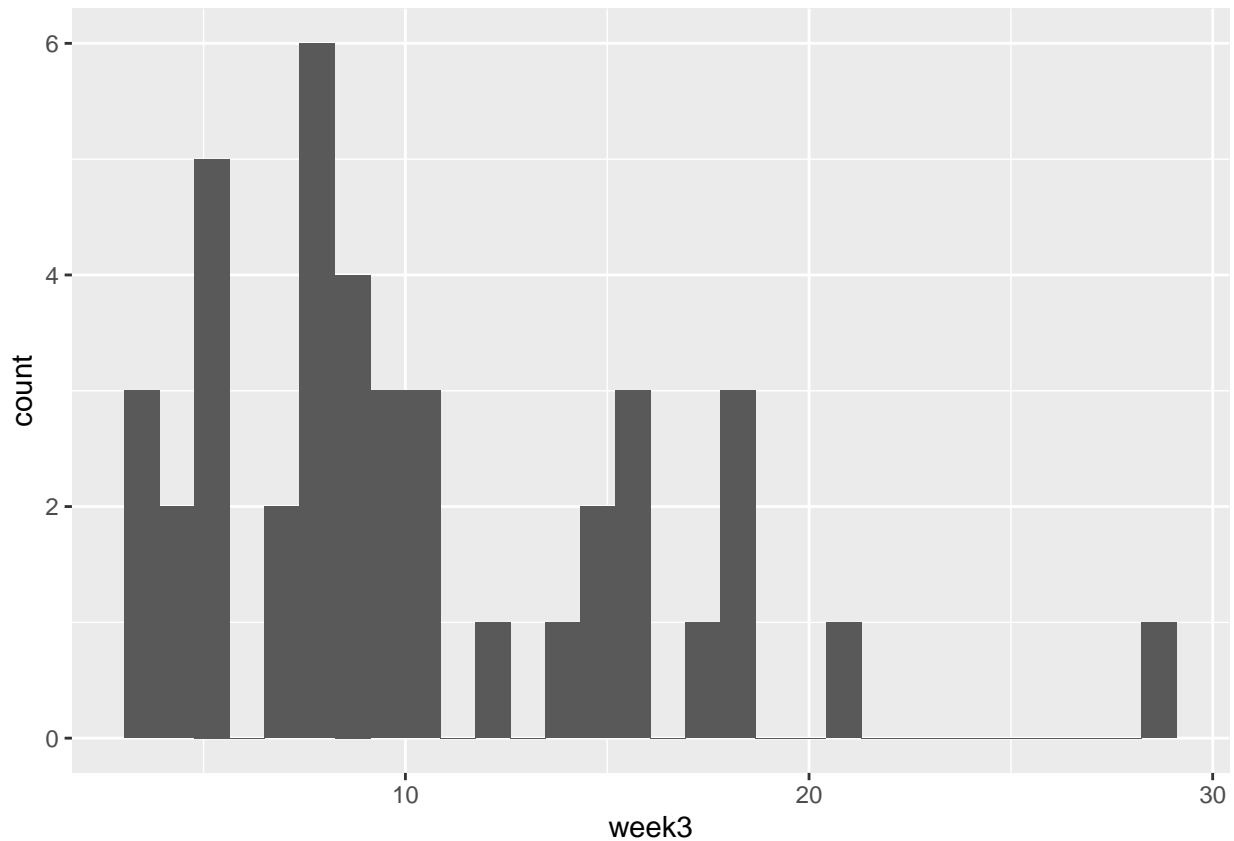
```
shapiro.test(Inulin_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Inulin_weekly$week3
## W = 0.91042, p-value = 0.003415
```

```
ggplot(Inulin_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.003415, kewed to the right -> no
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 7 rows containing non-finite values (stat_bin).
```



```
# Equal Variance
var.test(x = Inulin_weekly$week1,
         y = Inulin_weekly$week3,
         alternative = "two.sided") # p-value = 0.3398, variances are equal
```

```
##
## F test to compare two variances
##
## data: Inulin_weekly$week1 and Inulin_weekly$week3
## F = 1.3507, num df = 43, denom df = 40, p-value = 0.3398
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.7259875 2.4954729
## sample estimates:
## ratio of variances
##      1.350675
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
Inulin_weekly <- na.omit(Inulin_weekly)
wilcox.test(x = Inulin_weekly$week1,
            y = Inulin_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE)
```

```
##
## Wilcoxon signed rank test
```

```
##
## data: Inulin_weekly$week1 and Inulin_weekly$week3
## V = 333, p-value = 0.2978
## alternative hypothesis: true location shift is less than 0

# p-value = 0.2978, which is larger than 0.05 and we fail to reject the null. There is not sufficient evidence to reject the null.

# LOODAT
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "LOODAT") %>%
  summarize(sample_size = n()) # week1 n=16 & week3 n=16, not use the NA - WILCOX

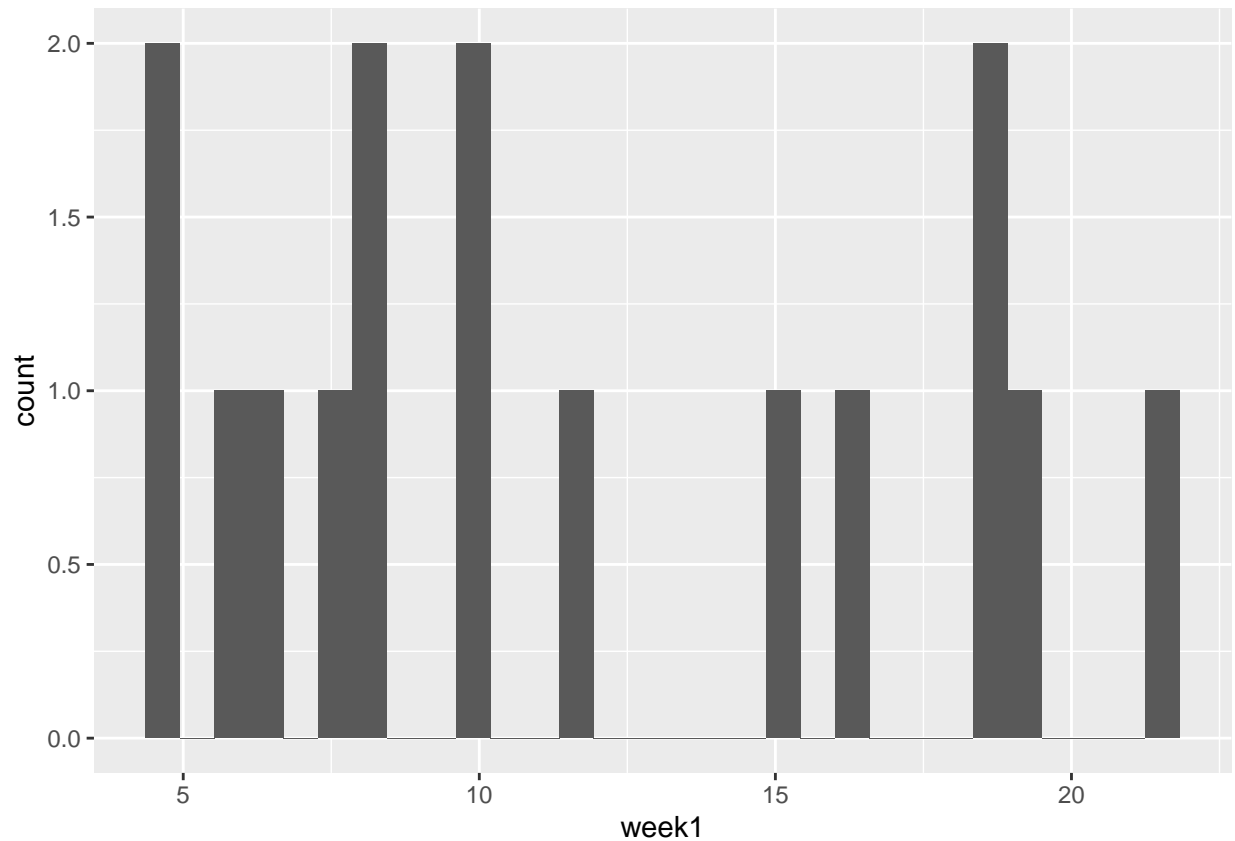
## # A tibble: 1 x 1
##   sample_size
##       <int>
## 1         16

# Normality
shapiro.test(LOODAT_weekly$week1)

##
## Shapiro-Wilk normality test
##
## data: LOODAT_weekly$week1
## W = 0.90517, p-value = 0.09729

ggplot(LOODAT_weekly, aes(x = week1)) + geom_histogram() # p-value = 0.09729, sort of bimodal and very skewed.

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

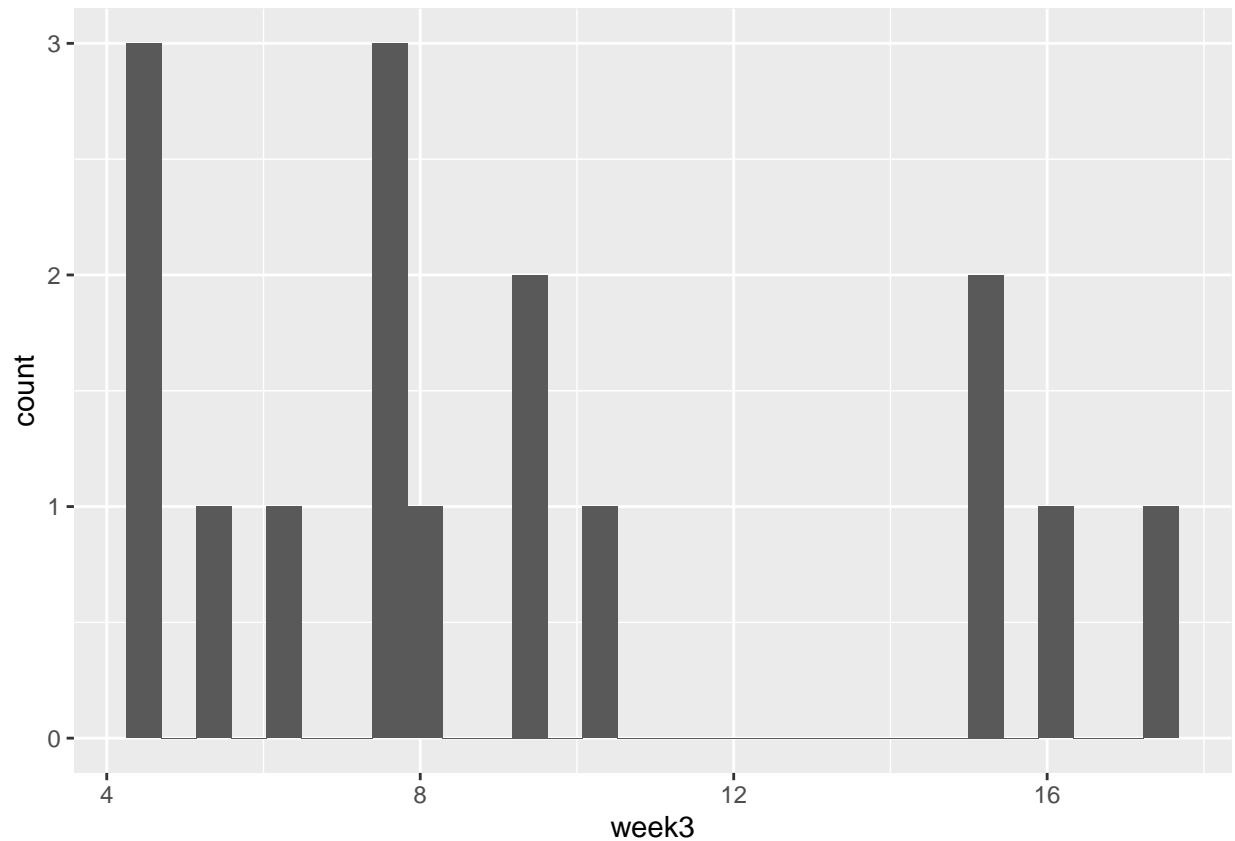


```
shapiro.test(LOODAT_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  LOODAT_weekly$week3
## W = 0.87837, p-value = 0.0366
```

```
ggplot(LOODAT_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.0366, sort of bimodal large spac
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
# Equal Variance
var.test(x = LOODAT_weekly$week1,
         y = LOODAT_weekly$week3,
         alternative = "two.sided") # p-value = 0.3428, variances are equal
```

```
##
## F test to compare two variances
##
## data: LOODAT_weekly$week1 and LOODAT_weekly$week3
## F = 1.6498, num df = 15, denom df = 15, p-value = 0.3428
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.576446 4.722000
## sample estimates:
## ratio of variances
## 1.649842
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
LOODAT_weekly <- na.omit(LOODAT_weekly)
wilcox.test(x = LOODAT_weekly$week1,
            y = LOODAT_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE)
```

```
##
## Wilcoxon signed rank test
```

```
##
## data: LOODAT_weekly$week1 and LOODAT_weekly$week3
## V = 94, p-value = 0.9123
## alternative hypothesis: true location shift is less than 0

# p-value = 0.9123, which is larger than 0.05 and we fail to reject the null. There is not sufficient evidence to reject the null.

# No Supplement
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "none") %>%
  summarize(sample_size = n()) # week 1 = 26, week 3 = 28

## # A tibble: 1 x 1
##   sample_size
##   <int>
## 1       29

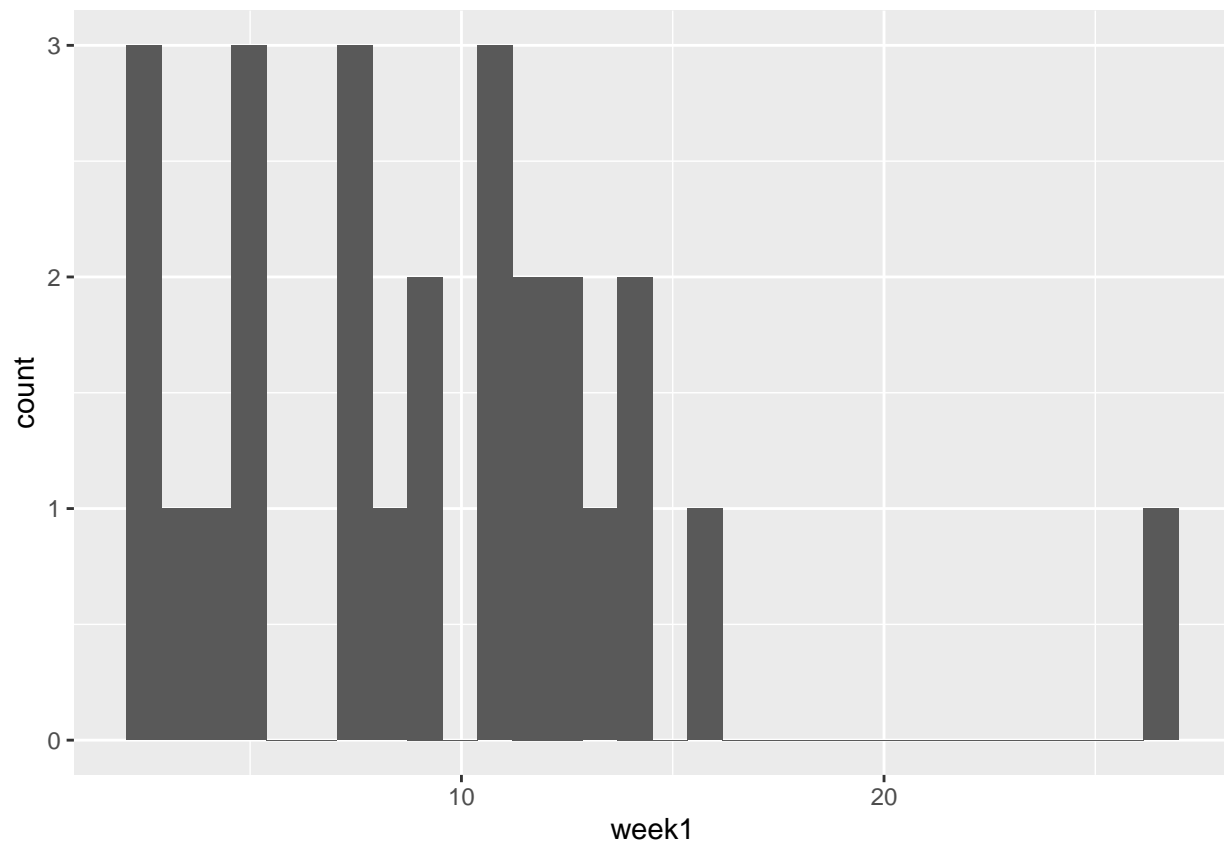
# Normality
shapiro.test(none_weekly$week1)

##
## Shapiro-Wilk normality test
##
## data: none_weekly$week1
## W = 0.89514, p-value = 0.01222

ggplot(none_weekly, aes(x = week1)) + geom_histogram() # p-value = 0.01, not normal

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 3 rows containing non-finite values (stat_bin).
```



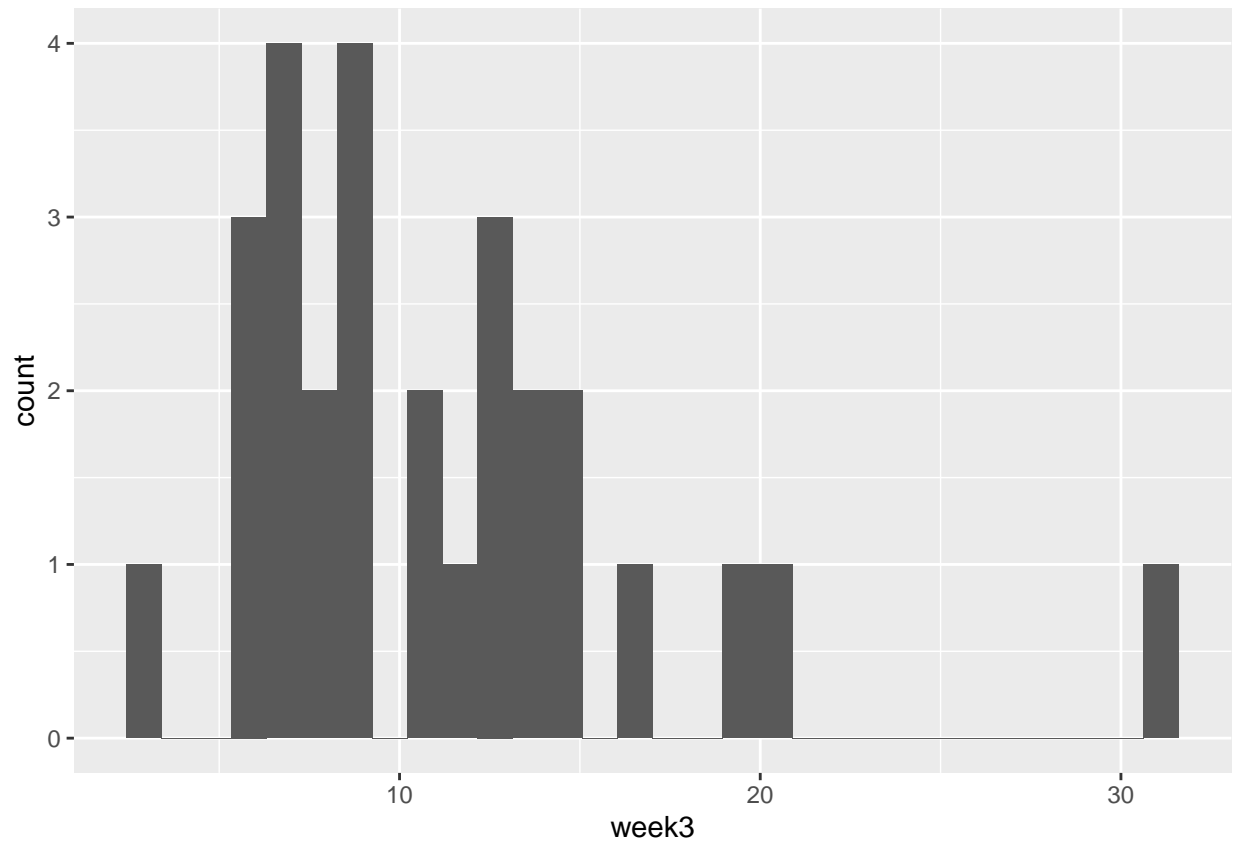
```
shapiro.test(none_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  none_weekly$week3
## W = 0.86441, p-value = 0.001852
```

```
ggplot(none_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.002, not normal
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



```
# Equal Variance
var.test(x = none_weekly$week1,
         y = none_weekly$week3,
         alternative = "two.sided") # p-value = 0.66, variances considered equal
```

```
##
## F test to compare two variances
##
## data: none_weekly$week1 and none_weekly$week3
## F = 0.83618, num df = 25, denom df = 27, p-value = 0.6559
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.3831176 1.8470162
## sample estimates:
## ratio of variances
## 0.8361819
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
wilcox.test(x = none_weekly$week1,
            y = none_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE, na.rm = TRUE)
```

```
##
## Wilcoxon signed rank test
##
```

```
## data: none_weekly$week1 and none_weekly$week3
## V = 117, p-value = 0.1152
## alternative hypothesis: true location shift is less than 0
```

```
# p-value = 0.12, which is larger than 0.05 and we fail to reject the null. there is not sufficient evi
```

```
# Psyllium
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "Psyllium") %>%
  summarize(sample_size = n()) # week 1 = 12, week 3 = 16
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1           18
```

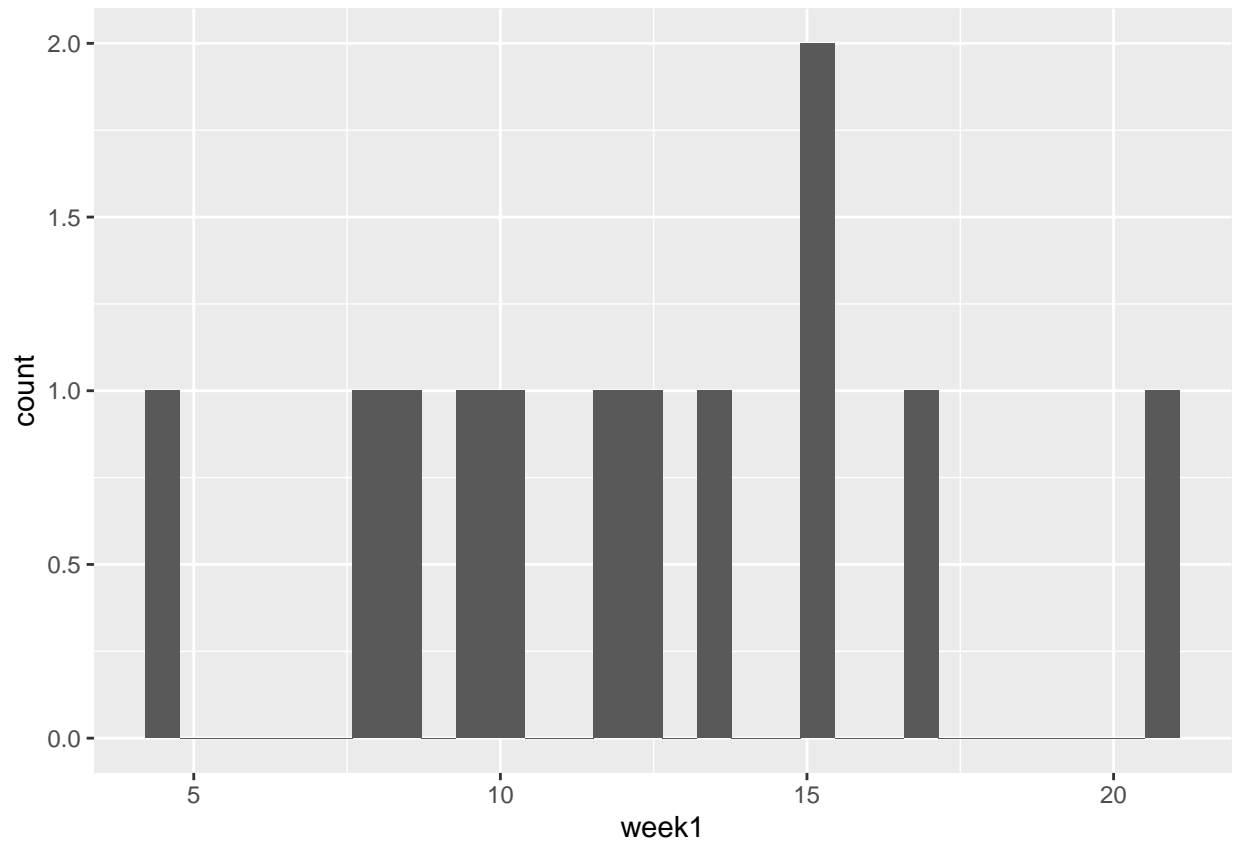
```
# Normality
shapiro.test(Psyllium_weekly$week1)
```

```
##
## Shapiro-Wilk normality test
##
## data: Psyllium_weekly$week1
## W = 0.99254, p-value = 1
```

```
ggplot(Psyllium_weekly, aes(x = week1)) + geom_histogram() # p-value = 1, normal
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 6 rows containing non-finite values (stat_bin).
```



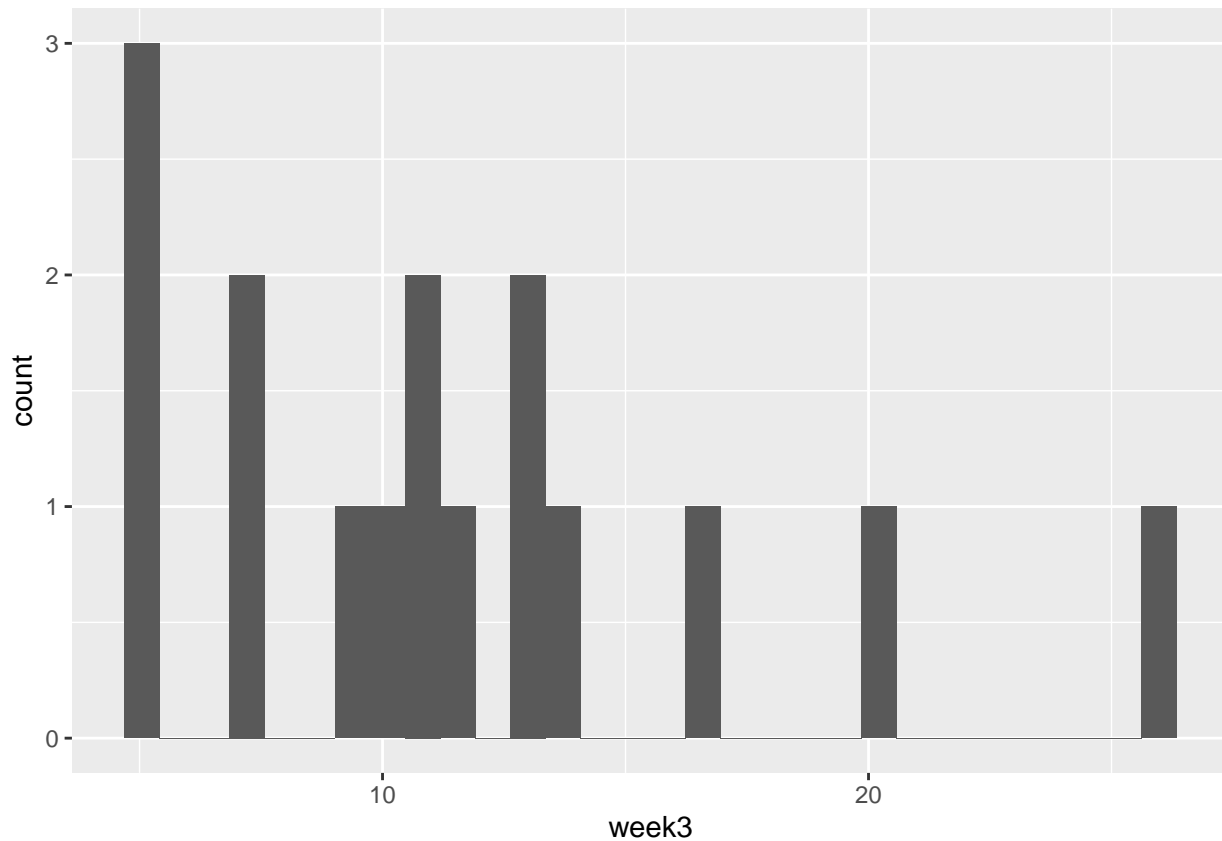
```
shapiro.test(Psyllium_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Psyllium_weekly$week3
## W = 0.91086, p-value = 0.1202
```

```
ggplot(Psyllium_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.12, normal
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 2 rows containing non-finite values (stat_bin).
```



```
# Equal Variance
var.test(x = Psyllium_weekly$week1,
         y = Psyllium_weekly$week3,
         alternative = "two.sided") # p-value = 0.41, variances considered equal
```

```
##
## F test to compare two variances
##
## data: Psyllium_weekly$week1 and Psyllium_weekly$week3
## F = 0.61003, num df = 11, denom df = 15, p-value = 0.4124
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.2028147 2.0313655
## sample estimates:
## ratio of variances
## 0.6100316
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
wilcox.test(x = Psyllium_weekly$week1,
            y = Psyllium_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE, na.rm = TRUE)
```

```
##
## Wilcoxon signed rank test
##
```

```
## data: Psyllium_weekly$week1 and Psyllium_weekly$week3
## V = 30, p-value = 0.6152
## alternative hypothesis: true location shift is less than 0
```

```
# p-value = 0.62, which is larger than 0.05 and we fail to reject the null. There is not sufficient evi
```

```
# Psyllium + BRMPS
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "Psyllium+BRMPS") %>%
  summarize(sample_size = n()) # week 1 = 14, week 3 = 15
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1           16
```

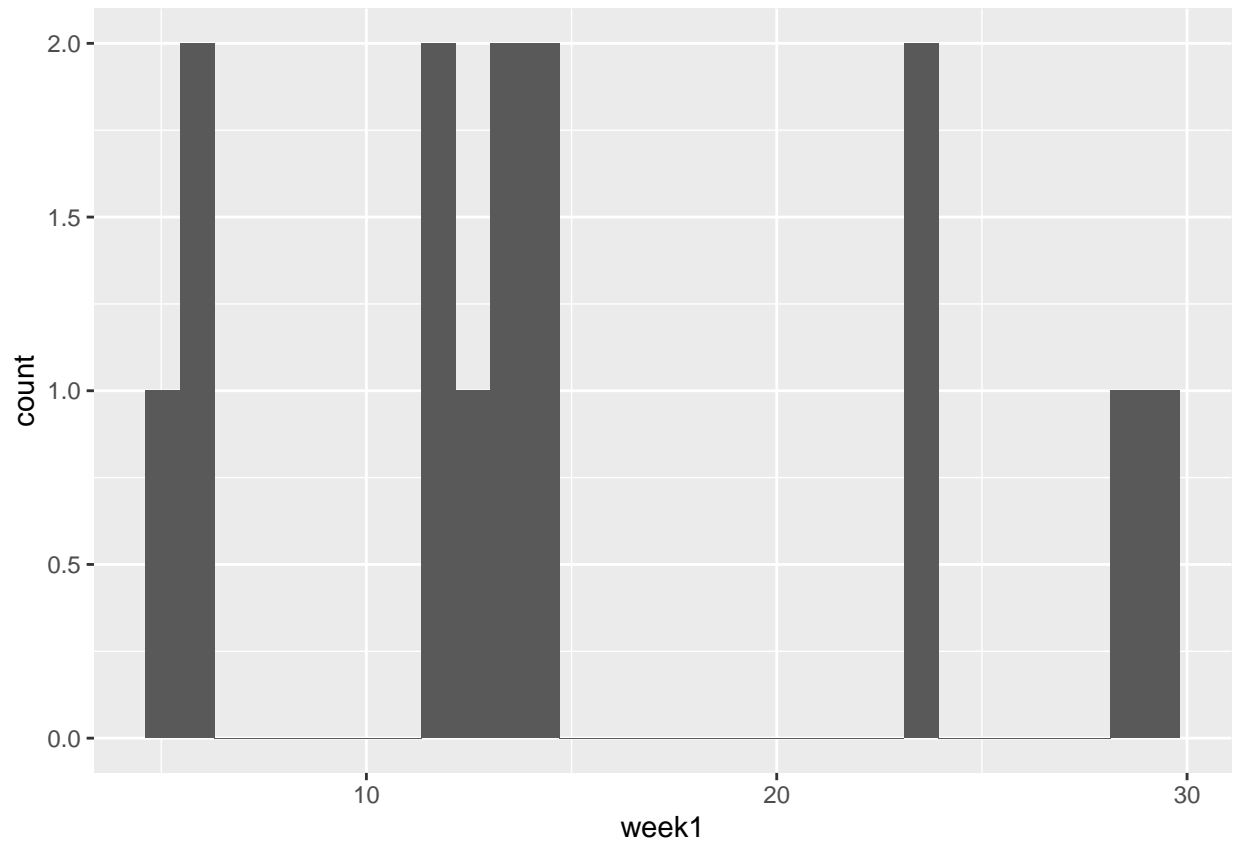
```
# Normality
shapiro.test(Psyllium_BRMPS_weekly$week1)
```

```
##
## Shapiro-Wilk normality test
##
## data: Psyllium_BRMPS_weekly$week1
## W = 0.88702, p-value = 0.07324
```

```
ggplot(Psyllium_BRMPS_weekly, aes(x = week1)) + geom_histogram() # p-value = 0.07, not normal
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 2 rows containing non-finite values (stat_bin).
```

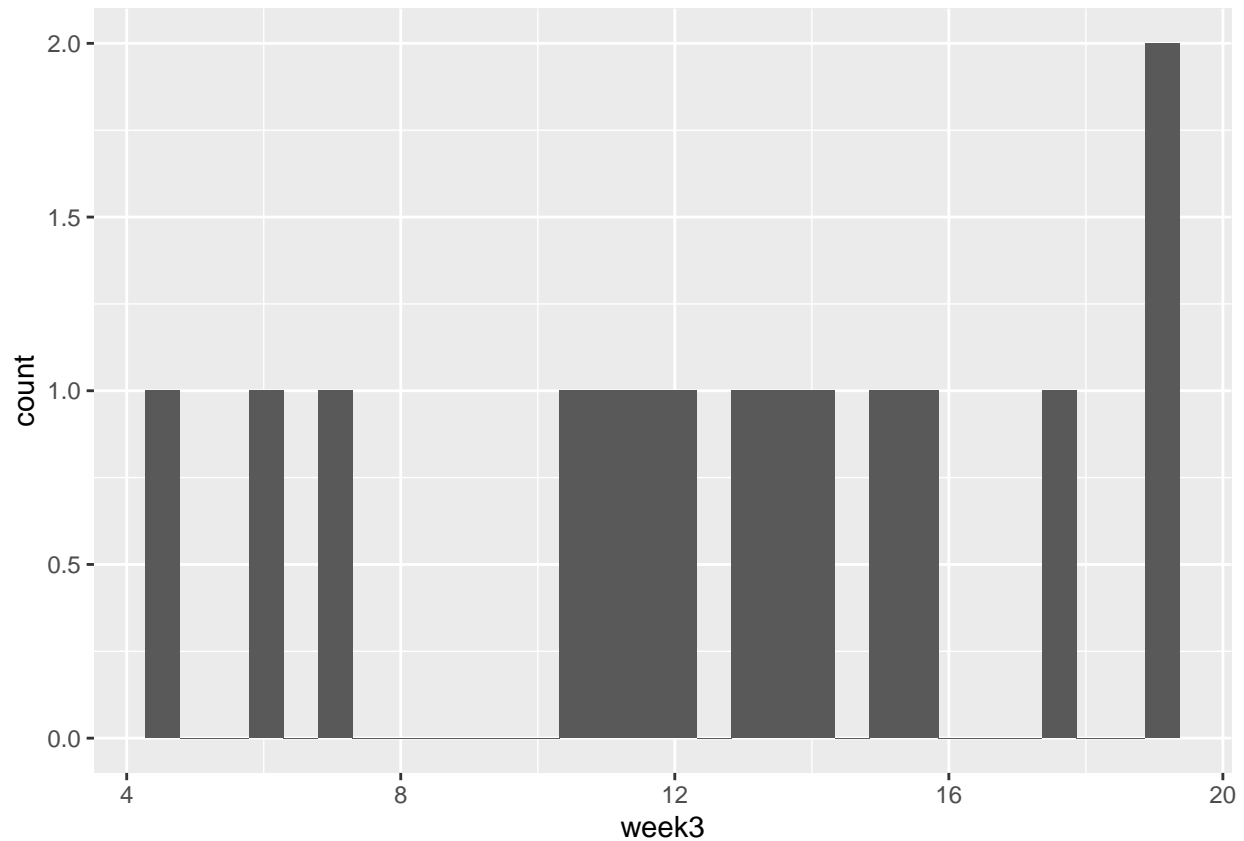
```
shapiro.test(Psyllium_BRMPs_weekly$week3)
```

```
##
## Shapiro-Wilk normality test
##
## data:  Psyllium_BRMPs_weekly$week3
## W = 0.95452, p-value = 0.5982
```

```
ggplot(Psyllium_BRMPs_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.60, normal
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



```
# Equal Variance
var.test(x = Psyllium_BRMPs_weekly$week1,
         y = Psyllium_BRMPs_weekly$week3,
         alternative = "two.sided") # p-value = 0.04, variances considered unequal
```

```
##
## F test to compare two variances
##
## data: Psyllium_BRMPs_weekly$week1 and Psyllium_BRMPs_weekly$week3
## F = 3.2565, num df = 13, denom df = 14, p-value = 0.03641
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  1.081202 10.035950
## sample estimates:
## ratio of variances
##      3.256465
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
wilcox.test(x = Psyllium_BRMPs_weekly$week1,
            y = Psyllium_BRMPs_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = FALSE, na.rm = TRUE)
```

```
##
## Wilcoxon signed rank test
##
```

```
## data: Psyllium_BRMPs_weekly$week1 and Psyllium_BRMPs_weekly$week3
## V = 62, p-value = 0.8781
## alternative hypothesis: true location shift is less than 0
```

p-value = 0.88, which is larger than 0.05 and we fail to reject the null. There is not sufficient evidence to reject the null.

```
# Transition HiMaize
# Sample Size
scfa_weekly %>%
  filter(supplement_consumed == "transition_HiMaize") %>%
  summarize(sample_size = n()) # week 1 = 27, week 3 = 27
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1           28
```

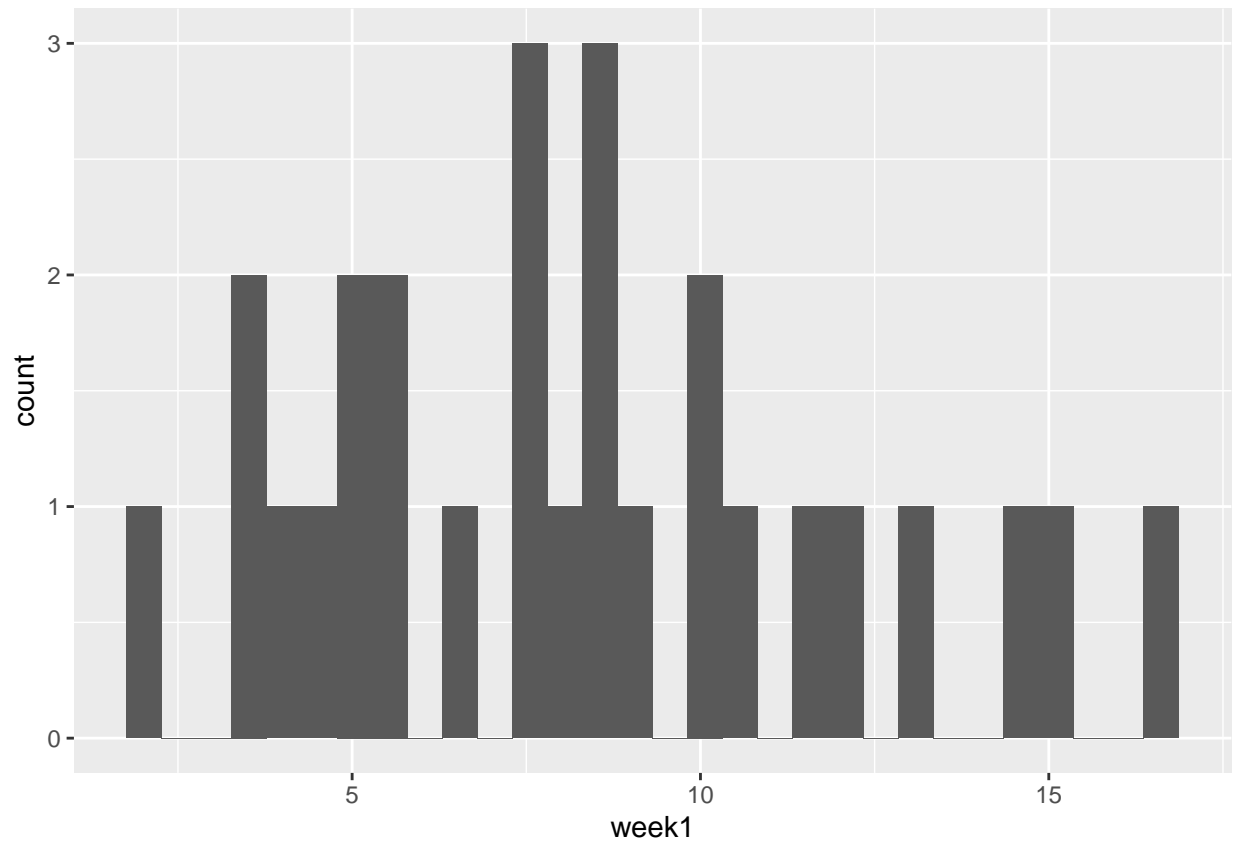
```
# Normality
shapiro.test(transition_HiMaize_weekly$week1)
```

```
##
## Shapiro-Wilk normality test
##
## data: transition_HiMaize_weekly$week1
## W = 0.9699, p-value = 0.5992
```

```
ggplot(transition_HiMaize_weekly, aes(x = week1)) + geom_histogram() # p-value = 0.60, normal
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



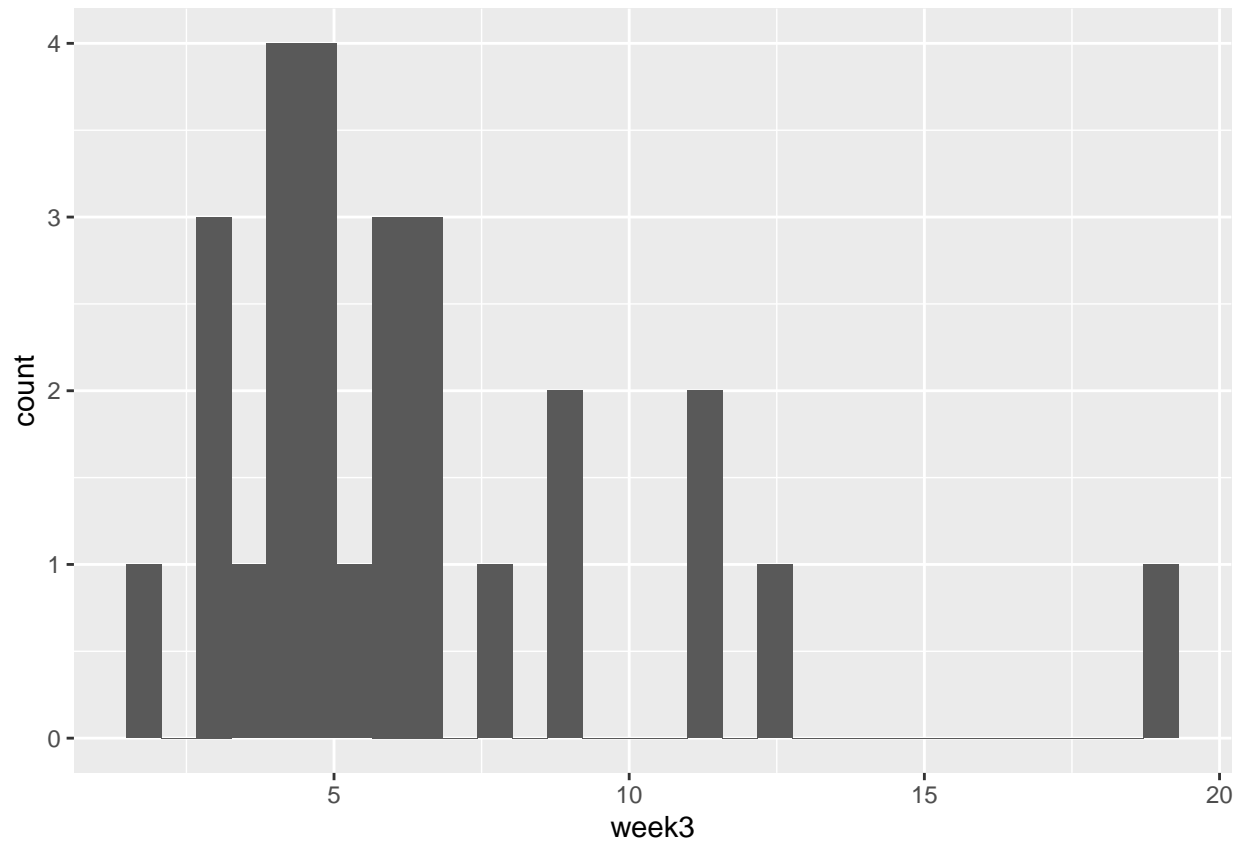
```
shapiro.test(transition_HiMaize_weekly$week3)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  transition_HiMaize_weekly$week3
## W = 0.82346, p-value = 0.0003589
```

```
ggplot(transition_HiMaize_weekly, aes(x = week3)) + geom_histogram() # p-value = 0.0004, not normal
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 1 rows containing non-finite values (stat_bin).
```



```
# Equal Variance
var.test(x = transition_HiMaize_weekly$week1,
         y = transition_HiMaize_weekly$week3,
         alternative = "two.sided") # p-value = 0.92, variances considered equal
```

```
##
## F test to compare two variances
##
## data: transition_HiMaize_weekly$week1 and transition_HiMaize_weekly$week3
## F = 1.0394, num df = 26, denom df = 26, p-value = 0.9222
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.4736998 2.2808544
## sample estimates:
## ratio of variances
## 1.039442
```

```
# Statistical Test (Wilcoxon because data is nonparametric)
wilcox.test(x = transition_HiMaize_weekly$week1,
            y = transition_HiMaize_weekly$week3,
            alternative = "less", paired = TRUE, var.equal = TRUE, na.rm = TRUE)
```

```
##
## Wilcoxon signed rank test
##
```

```
## data: transition_HiMaize_weekly$week1 and transition_HiMaize_weekly$week3
## V = 249, p-value = 0.9703
## alternative hypothesis: true location shift is less than 0
```

p-value = 0.97, which is larger than 0.05 and we fail to reject the null. There is not sufficient evidence to reject the null.

Plots!

Correlation plots

delta changes we first calculate the change of produced proprionates in the population who consumes t

```
HiMaize_weekly_delta <- HiMaize_weekly %>%
mutate(delta_propionate = week3 - week1) %>%
gather(key=study_week, value= proprionate_value, week1:week3)

write_delim(HiMaize_weekly_delta, path = "~/Documents/Bio201_Project/curated_data/long_priop_delta.txt")

long_priop_delta_table <- HiMaize_weekly_delta %>%
  mutate(sample_id = paste(participant_id, study_week, sep="_")) %>% #cobine two columns to make new co
  column_to_rownames(var = "sample_id") %>%
sample_data(.)
```

seq var info

```
all_data_indv <- read_delim("raw_data/species_avg_shared.txt",
                           delim = "\t", escape_double = FALSE,
                           trim_ws = TRUE,
                           col_types = cols()) %>%
```

```
  rename_all(tolower) %>%
  filter(semester == "Winter2017" | semester == "Fall2015") %>%
  select(-semester) %>%
  mutate(sample_id = paste(participant_id, study_week, sep="_")) %>%
  distinct(., sample_id, .keep_all = TRUE) %>%
  column_to_rownames(var = "sample_id") %>%
  select(-participant_id, -study_week) %>%
  as.matrix() %>%
  otu_table(., taxa_are_rows = FALSE)
```

merge all three data frames into phyloseq object

```
physq_obj <- phyloseq(all_data_indv, long_priop_delta_table)
physq_obj
```

```
## phyloseq-class experiment-level object
```

```
## otu_table() OTU Table: [ 1160 taxa and 49 samples ]
```

```
## sample_data() Sample Data: [ 49 samples by 8 sample variables ]
```

#long format of species and people data

```
sample_data_long <- psmelt(physq_obj) %>%
rename_all(tolower) %>%
rename(tax_name = otu ) %>%
```

```

    filter(abundance > 0)

generator_info<-read_delim(file = "~/Documents/Bio201_Project/raw_data/tax_scfa.txt",
                          delim = "\t", col_names = TRUE, trim_ws = TRUE,
                          na = c("", "NA")) %>%
  rename_all(tolower) %>%
  select(tax_name, category, phylum)

```

```

## Parsed with column specification:
## cols(
##   participant_id = col_character(),
##   study_week = col_character(),
##   frequency = col_character(),
##   semester = col_character(),
##   supplement_consumed = col_character(),
##   quantity_compliant = col_character(),
##   acetate_mean = col_double(),
##   butyrate_mean = col_double(),
##   propionate_mean = col_double(),
##   totalscfa_mean = col_double(),
##   tax_id = col_double(),
##   relative_abundance = col_double(),
##   phylum = col_character(),
##   tax_name = col_character(),
##   category = col_character()
## )

```

```

generator_info$tax_name <- tolower(generator_info$tax_name)
sample_data_long_cut <- sample_data_long %>%
  select(sample, tax_name, abundance,delta_propionate, proprionate_value,semester)
relative_sample <- inner_join(sample_data_long_cut,generator_info,
                              by = ("tax_name"))
relative_sample <- unique(relative_sample [,c("sample","tax_name", "category", "abundance", "delta_propionate")])

```

tidy data for correlation

```

# week 3 data with mached metabolic identities
grep_week3<- relative_sample[grep("week3", relative_sample$sample), ]

#relative abundance of proprionate generators in week 3 in participants who consume himaze
prop_gen_abd <- grep_week3 %>%
  filter (category == "propionate_gen") %>%
  group_by(sample) %>%
  summarize(prop_abundance = sum(abundance))
delta_propionate <- grep_week3 %>%
  select(sample,delta_propionate,proprionate_value)%>%
  distinct(., sample, .keep_all = TRUE)
graph_data<- inner_join(prop_gen_abd,delta_propionate,
                        by = ("sample"))

#propionate generator abundance for group1

```

```

grep_week1<- relative_sample[grepl("week1", relative_sample$sample), ]
prop_gen_abd_1 <- grep_week1 %>%
  filter (category == "propionate_gen") %>%
  select(-tax_name) %>%
  group_by(sample) %>%
  summarize(prop_abundance = sum(abundance))
gen_abd_1<- prop_gen_abd_1 %>%
  mutate(Participant_ID = paste(sample)) %>% #copy column
  separate(Participant_ID,
            sep = "_",
            into = c("Participant_ID", "study_week")) %>%
  select(-sample,-study_week) %>%
  rename(prop_abundance_wk1 = prop_abundance)

delta_propionate <- grep_week1 %>%
  select(sample,delta_propionate,propionate_value)%>%
  distinct(., sample, .keep_all = TRUE)
graph_data_1<- inner_join(prop_gen_abd_1,delta_propionate,
                          by = ("sample"))
prop_gen_abd <- grep_week3 %>%
  filter (category == "propionate_gen") %>%
  group_by(sample) %>%
  summarize(prop_abundance = sum(abundance))
gen_abd_3 <- prop_gen_abd %>%
  mutate(Participant_ID = paste(sample)) %>%
  separate(Participant_ID,
            sep = "_",
            into = c("Participant_ID", "study_week")) %>%
  select(-sample,-study_week) %>%
  rename(prop_abundance_wk3 = prop_abundance)
prop_gen_abd_all <- inner_join(gen_abd_1,gen_abd_3,
                              by = c("Participant_ID")) %>%
  mutate(delta_abundance = prop_abundance_wk3 - prop_abundance_wk1)

```

Relationship with propionate generators

```

graph_data_1 <- graph_data%>%
  filter(delta_propionate >0)
mean(graph_data_1$delta_propionate) #average of 4.5 % increase.

```

```
## [1] 4.495
```

```

graph_data_6 <- graph_data %>%
  filter(delta_propionate < 0)
mean(graph_data_6$delta_propionate)

```

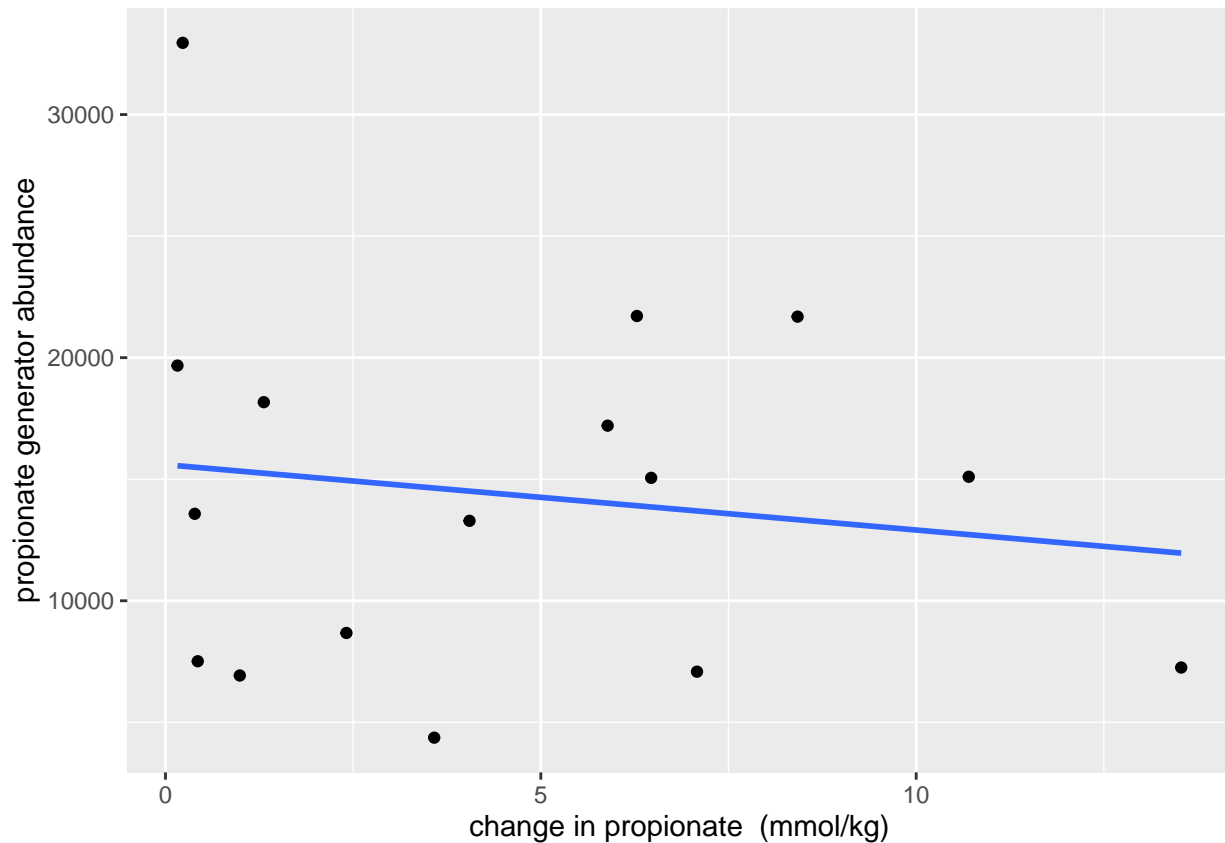
```
## [1] -2.612
```



```

# plot to determine whether there is a relationship between propionate abundance and positive delta_prop
plot_prop_delta<- graph_data_1 %>%
  ggplot(aes(x = delta_propionate,
             y = prop_abundance)) +
  geom_point() + #puts data points to match x and y coordinates
  geom_smooth(method = "lm", #used to create a linear best fit line
             se = FALSE) + #hides confidence interval around line
  xlab("change in propionate (mmol/kg)") +
  ylab("propionate generator abundance")
plot_prop_delta

```



```

#correlation test
m_delta <- graph_data_1 %>%
  lm(delta_propionate ~ prop_abundance, data = .) #test relationship
summary(m_delta) #view results

```

```

##
## Call:
## lm(formula = delta_propionate ~ prop_abundance, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.611  -3.143  -1.122   2.112   8.468
##
## Coefficients:

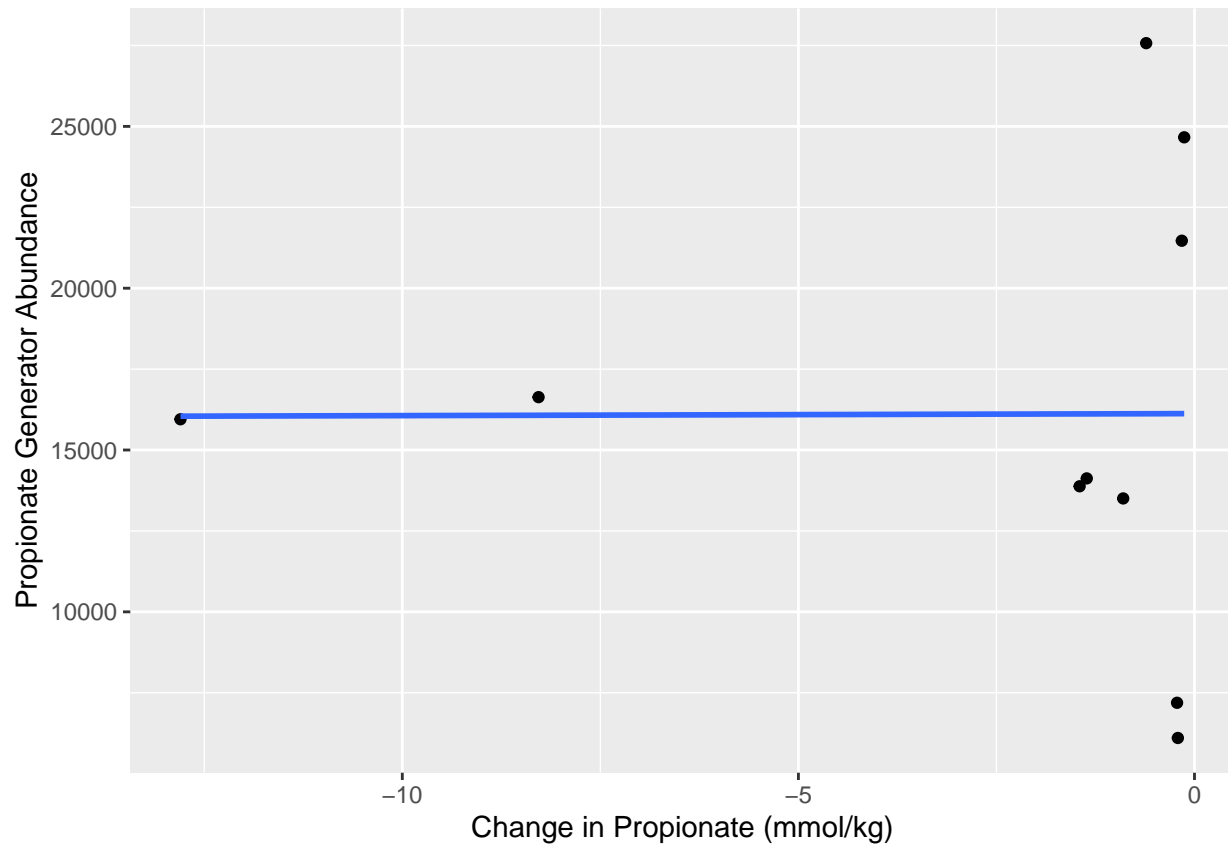
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.638e+00  2.316e+00   2.434  0.0289 *
## prop_abundance -7.941e-05  1.437e-04  -0.553  0.5892
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.176 on 14 degrees of freedom
## Multiple R-squared:  0.02135,    Adjusted R-squared:  -0.04855
## F-statistic: 0.3055 on 1 and 14 DF,  p-value: 0.5892
```

*#the p value is 0.5892, larger than 0.05, which means we can't use the linear model to interpret the
#is there a relationship between propionate generator abundance and the propionate value in group who s*

plot to determine whether there is a relationship between propionate abundance and negative delta_prop

```
prop_plot <- graph_data_6 %>%
  ggplot(aes(x = delta_propionate,
             y = prop_abundance)) +
  geom_point() +
  geom_smooth(method = "lm",
             se = FALSE) +
  xlab("Change in Propionate (mmol/kg)") +
  ylab("Propionate Generator Abundance")
prop_plot
```



```

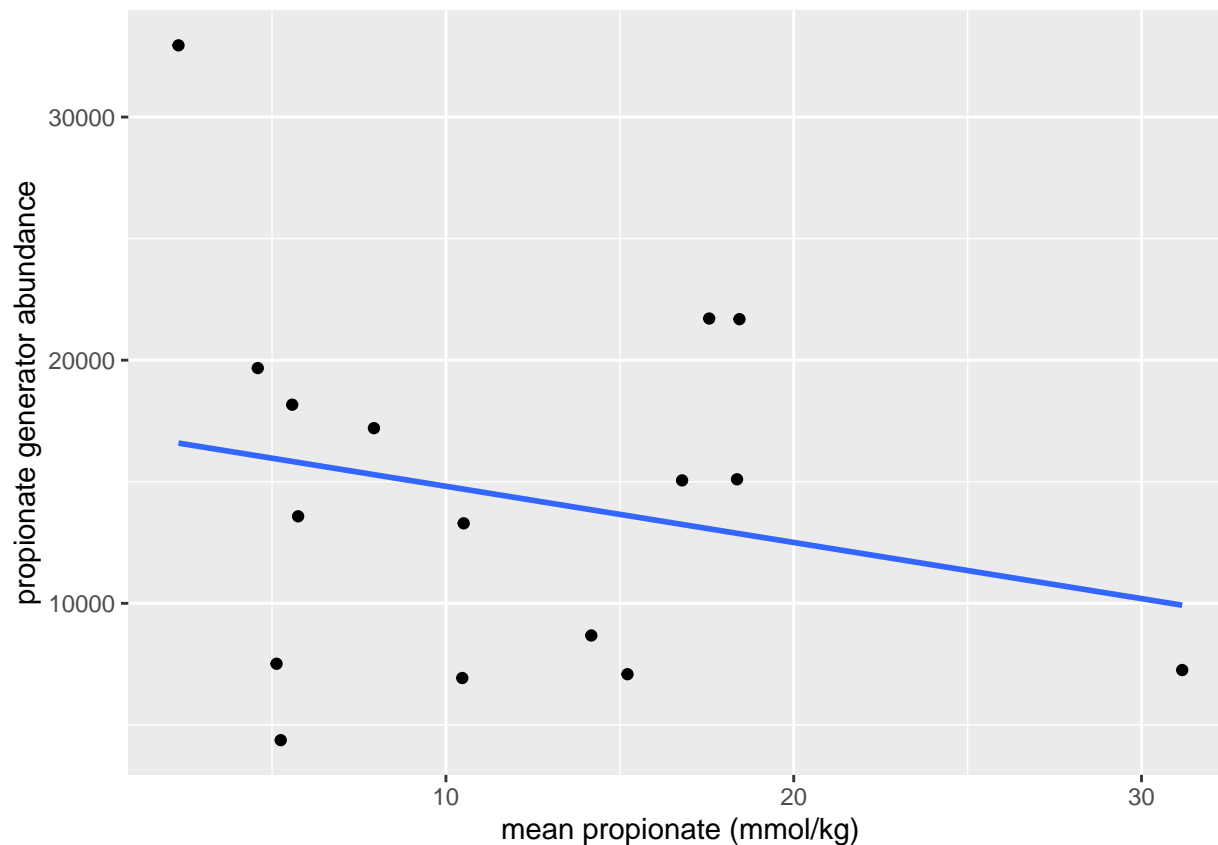
# Correlation test
m_delta <- graph_data_6 %>%
  lm(delta_propionate ~ prop_abundance, data = .)
summary(m_delta)

##
## Call:
## lm(formula = delta_propionate ~ prop_abundance, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.188   1.190   1.846   2.423   2.461
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.651e+00  3.864e+00  -0.686   0.512
## prop_abundance  2.443e-06  2.222e-04   0.011   0.991
##
## Residual standard error: 4.602 on 8 degrees of freedom
## Multiple R-squared:  1.511e-05, Adjusted R-squared:  -0.125
## F-statistic: 0.0001209 on 1 and 8 DF, p-value: 0.9915

# p-value = 0.9915, there is not enough evidence to conclude a linear relationship between change in pr

plot_prop_value<- graph_data_1 %>%
  ggplot(aes(x = proprionate_value,
             y = prop_abundance)) +
  geom_point() + #puts data points to match x and y coordinates
  geom_smooth(method = "lm", #used to create a linear best fit line
             se = FALSE) + #hides confidence interval around line
  xlab("mean propionate (mmol/kg)") +
  ylab("propionate generator abundance")
plot_prop_value

```

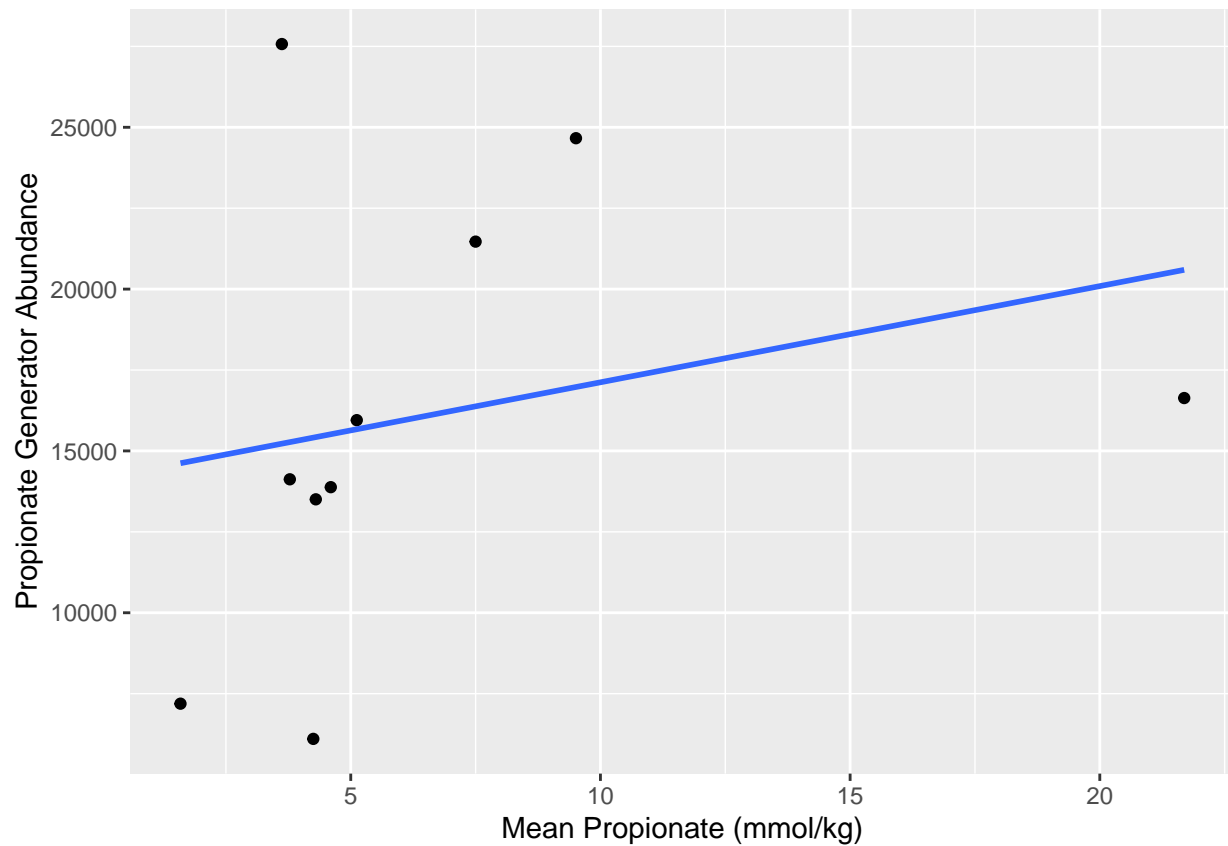


```
#correlation test
m_value <- graph_data_1 %>%
  lm(propionate_value ~ prop_abundance, data = .) #test relationship
summary(m_value) #view results
```

```
##
## Call:
## lm(formula = propionate_value ~ prop_abundance, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.939 -5.518 -2.348  5.516 17.660
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.2185304  4.2301955   3.598  0.00291 **
## prop_abundance -0.0002356  0.0002624  -0.898  0.38454
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.626 on 14 degrees of freedom
## Multiple R-squared:  0.05443,    Adjusted R-squared:  -0.01311
## F-statistic: 0.8058 on 1 and 14 DF,  p-value: 0.3845
```

#the p value is 0.536, larger than 0.05, which means we can't use the linear model to interpret the

```
prop_conc_plot <- graph_data_6 %>%  
  ggplot(aes(x = proprionate_value,  
             y = prop_abundance)) +  
  geom_point() +  
  geom_smooth(method = "lm",  
             se = FALSE) +  
  xlab("Mean Propionate (mmol/kg)") +  
  ylab("Propionate Generator Abundance")  
prop_conc_plot
```



```
# Correlation test  
m_conc <- graph_data_6 %>%  
  lm(proprionate_value ~ prop_abundance, data = .)  
summary(m_conc)
```

```
##  
## Call:  
## lm(formula = proprionate_value ~ prop_abundance, data = .)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -5.3266 -2.2472 -1.4913 -0.2196 14.9863   
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.2928283  4.9485054   0.665   0.524
## prop_abundance 0.0002051  0.0002846   0.721   0.492
##
## Residual standard error: 5.894 on 8 degrees of freedom
## Multiple R-squared:  0.06095,    Adjusted R-squared:  -0.05643
## F-statistic: 0.5192 on 1 and 8 DF,  p-value: 0.4917
```

p-value = 0.4917, there is not enough evidence to conclude a linear relationship between mean propion

```
save_plot(filename = "figures/abundance_vs_mean.pdf",
          plot= plot_prop_delta,
          nrow = 1, ncol = 2,
          base_aspect_ratio = 1.1)

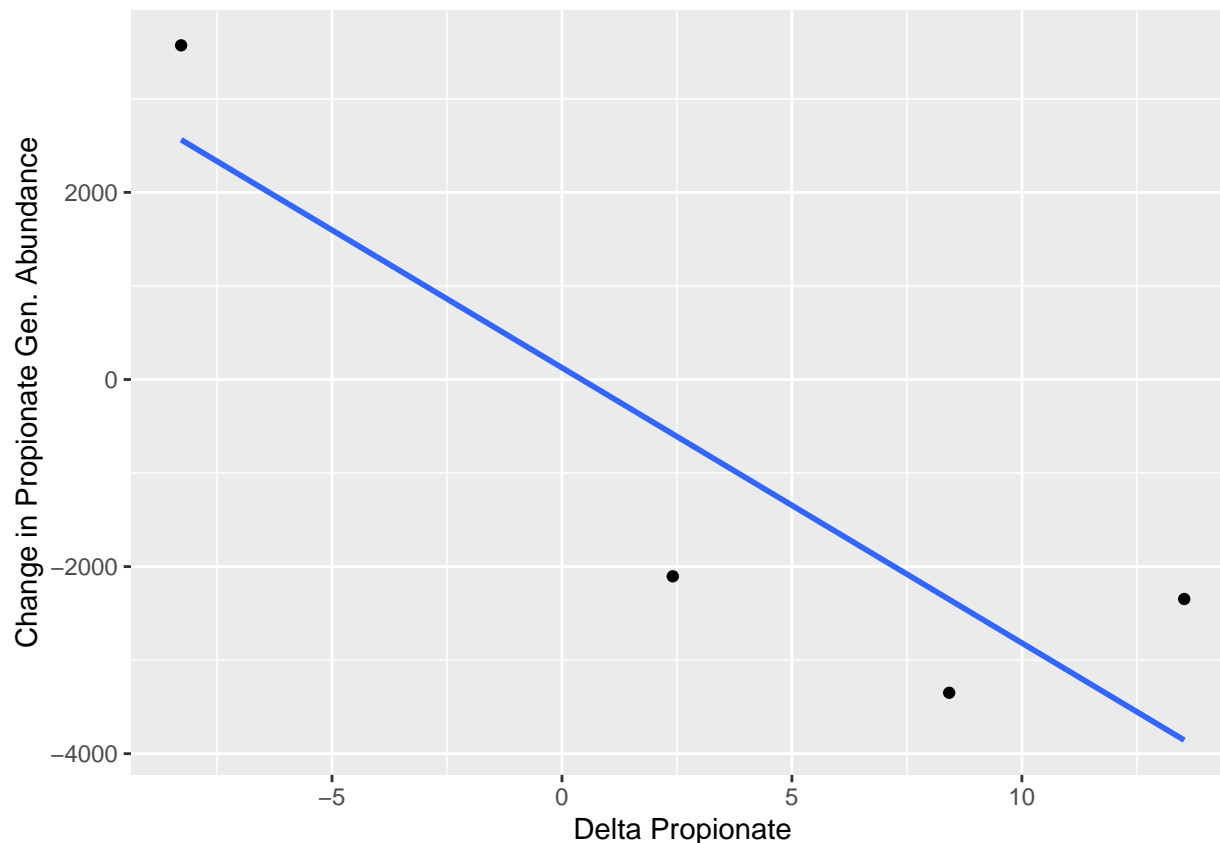
save_plot(filename = "figures/abundance_vs_delta.pdf",
          plot= plot_prop_value,
          nrow = 1, ncol = 2,
          base_aspect_ratio = 1.1)
```

Correlation separete by semester

```
# correlation: difference in semster [since different semester have different pariticipants with differ

sample_data_semster <- sample_data_long %>%
  select(semester, participant_id, delta_propionate, supplement_consumed)
sample_data_semster <- unique(sample_data_semster [,c("participant_id", "delta_propionate", "supplemen
prop_gen_abd_all <- prop_gen_abd_all %>%
  rename_all (tolower)
# data with delta abundance, semster, and delta propionate
merged_data <- inner_join(prop_gen_abd_all, sample_data_semster, by = c("participant_id"))
write_delim(merged_data, path = "~/Documents/Bio201_Project/curated_data/merged_data.txt", delim = "\t")

# Fall 2015
fall_sem_plot <- merged_data %>%
  filter(semester == "Fall2015") %>%
  ggplot(aes(x = delta_propionate,
            y = delta_abundance)) +
  geom_point() +
  geom_smooth(method = "lm",
            se = FALSE) +
  xlab("Delta Propionate") +
  ylab("Change in Propionate Gen. Abundance")
fall_sem_plot
```



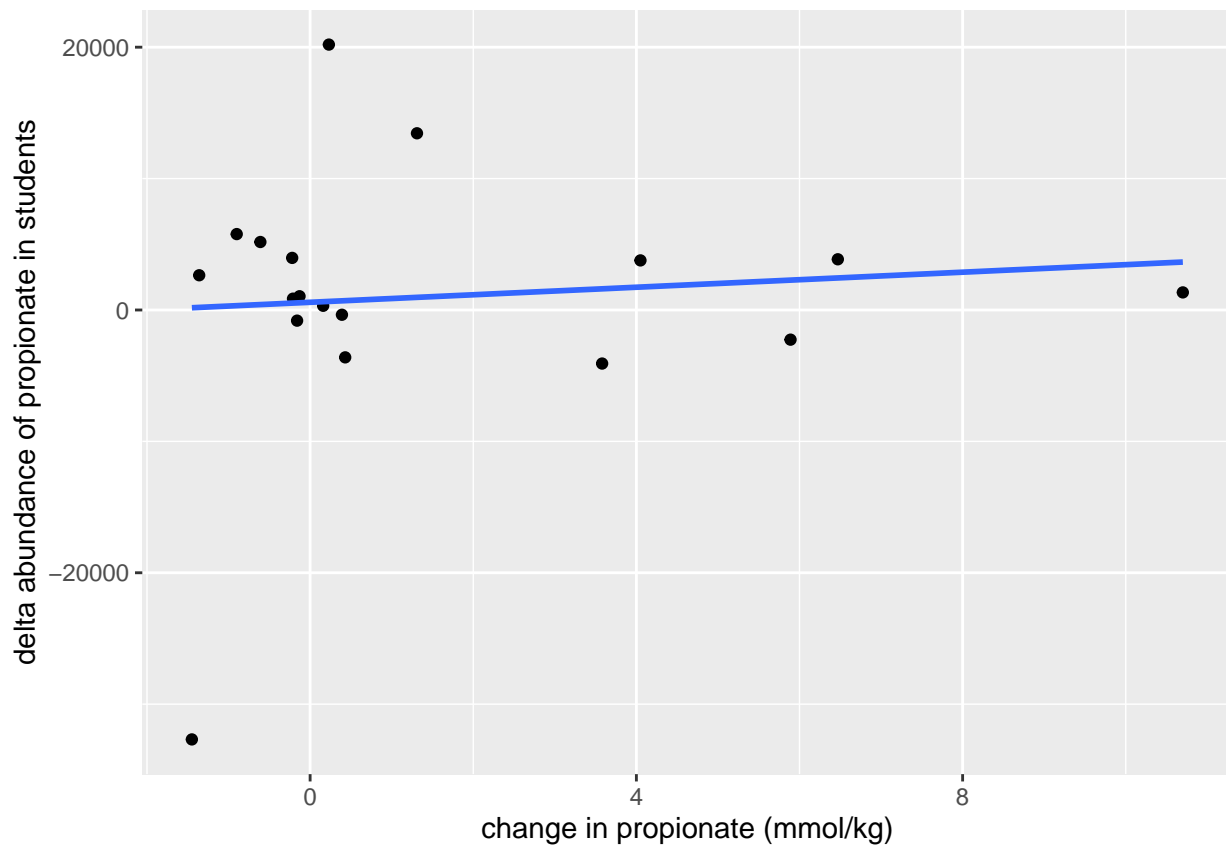
```
save_plot(filename = "figures/fall2015_prop_abund.pdf",
  plot= fall_sem_plot,
  nrow = 2, ncol = 3,
  base_aspect_ratio = 1.1)
m_fall <- merged_data %>%
  filter(semester == "Fall2015") %>%
  lm(delta_propionate ~ delta_abundance, data = .)
summary(m_fall) # p-value = 0.1193, there is not enough evidence to conclude a linear relationship betw
```

```
##
## Call:
## lm(formula = delta_propionate ~ delta_abundance, data = .)
##
## Residuals:
##      1      2      3      4
## -0.09911 -1.64170  6.11093 -4.37012
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.234317   2.918099   0.423   0.713
## delta_abundance -0.002636   0.001002  -2.630   0.119
##
## Residual standard error: 5.438 on 2 degrees of freedom
## Multiple R-squared:  0.7757, Adjusted R-squared:  0.6635
## F-statistic: 6.916 on 1 and 2 DF, p-value: 0.1193
```

```

# Winter 2017
sample_data_semster <- sample_data_long %>%
  select(semester, participant_id, delta_propionate, supplement_consumed)
sample_data_semster <- unique(sample_data_semster [,c("participant_id", "delta_propionate", "supplement_consumed")])
prop_gen_abd_all <- prop_gen_abd_all %>%
  rename_all(tolower)
# data with delta abundance, semester, and delta propionate
merged_data <- inner_join(prop_gen_abd_all, sample_data_semster, by = c("participant_id"))
write_delim(merged_data, path = "~/Documents/Bio201_Project/curated_data/merged_data.txt", delim = "\t")
graph_3 <- merged_data %>%
  filter(semester == "Winter2017")
plot_winter2017 <- graph_3 %>%
  ggplot(aes(x = delta_propionate,
             y = delta_abundance)) +
  geom_point() + #puts data points to match x and y coordinates
  geom_smooth(method = "lm", #used to create a linear best fit line
             se = FALSE) + #hides confidence interval around line
  xlab("change in propionate (mmol/kg)") +
  ylab("delta abundance of propionate in students")
plot_winter2017

```



```

save_plot(filename = "figures/w2017_deltaa_deltap.pdf",
  plot= plot_winter2017,
  nrow = 1, ncol = 1,
  base_aspect_ratio = 1.1)

```

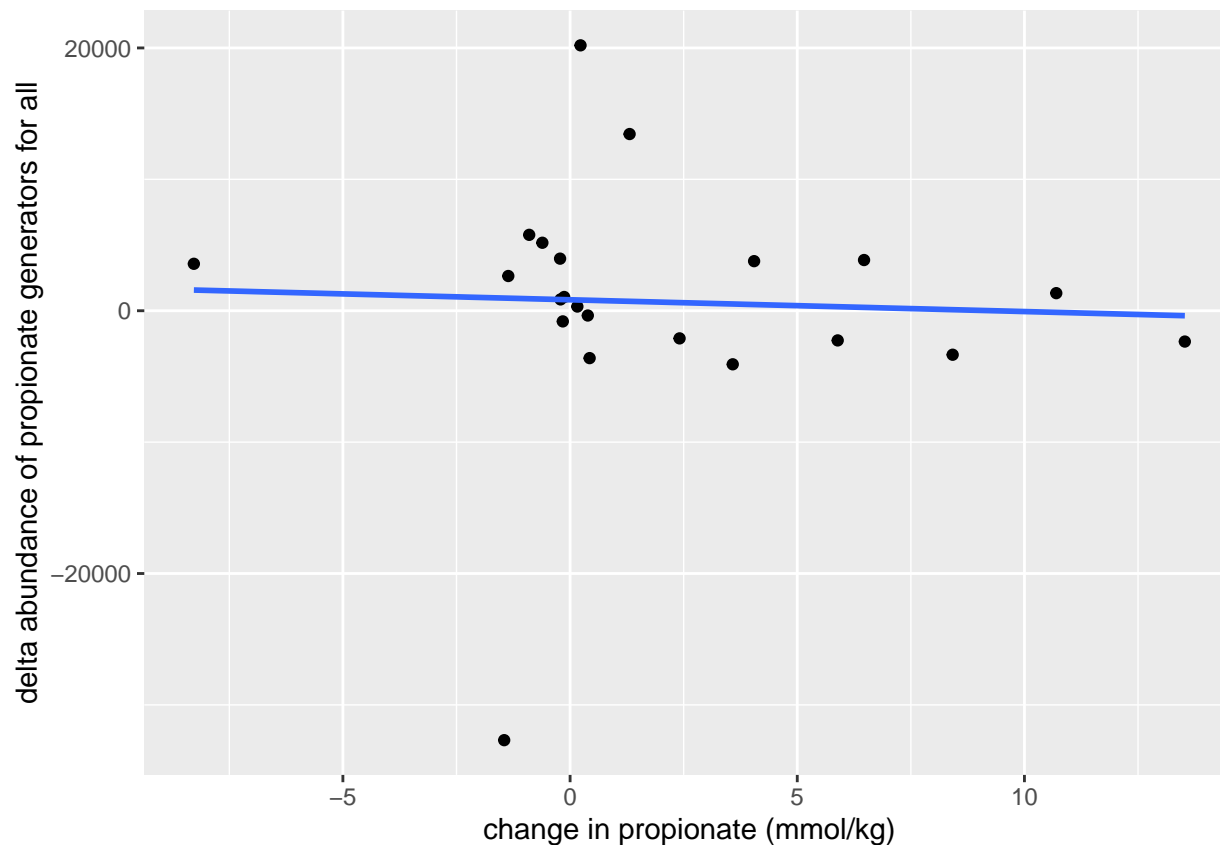


```
#correlation test
m_winter <- graph_3 %>%
  lm(delta_propionate ~ delta_abundance, data = .) #test relationship
summary(m_winter) #view results
```

```
##
## Call:
## lm(formula = delta_propionate ~ delta_abundance, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.972 -1.891 -1.527  1.469  9.126
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.535e+00  7.993e-01   1.920  0.0729 .
## delta_abundance 2.938e-05  7.974e-05   0.368  0.7174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.373 on 16 degrees of freedom
## Multiple R-squared:  0.008413,    Adjusted R-squared:  -0.05356
## F-statistic: 0.1358 on 1 and 16 DF,  p-value: 0.7174
```

*#the p value is 0.7174, larger than 0.05, which means we can't use the linear model to interpret the
(correlation: difference in abundance vs delta propionate (all))*

```
plot_delta <- merged_data %>%
  ggplot(aes(x = delta_propionate,
             y = delta_abundance)) +
  geom_point() + #puts data points to match x and y coordinates
  geom_smooth(method = "lm", #used to create a linear best fit line
             se = FALSE) + #hides confidence interval around line
  xlab("change in propionate (mmol/kg)") +
  ylab("delta abundance of propionate generators for all")
plot_delta
```



```
save_plot(filename = "figures/deltaa_deltap_all.pdf",
          plot= plot_delta,
          nrow = 2, ncol = 2,
          base_aspect_ratio = 1.1)
#correlation test
m_delta <- merged_data %>%
  lm(delta_propionate ~ delta_abundance, data = .) #test relationship
summary(m_delta) #view results
```

```
##
## Call:
## lm(formula = delta_propionate ~ delta_abundance, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.225  -2.214  -1.662   1.948  11.450
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.026e+00  1.032e+00   1.964  0.0636 .
## delta_abundance -2.282e-05  1.128e-04  -0.202  0.8417
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.827 on 20 degrees of freedom
```

```
## Multiple R-squared:  0.002044,   Adjusted R-squared:  -0.04785
## F-statistic: 0.04096 on 1 and 20 DF,  p-value: 0.8417
```

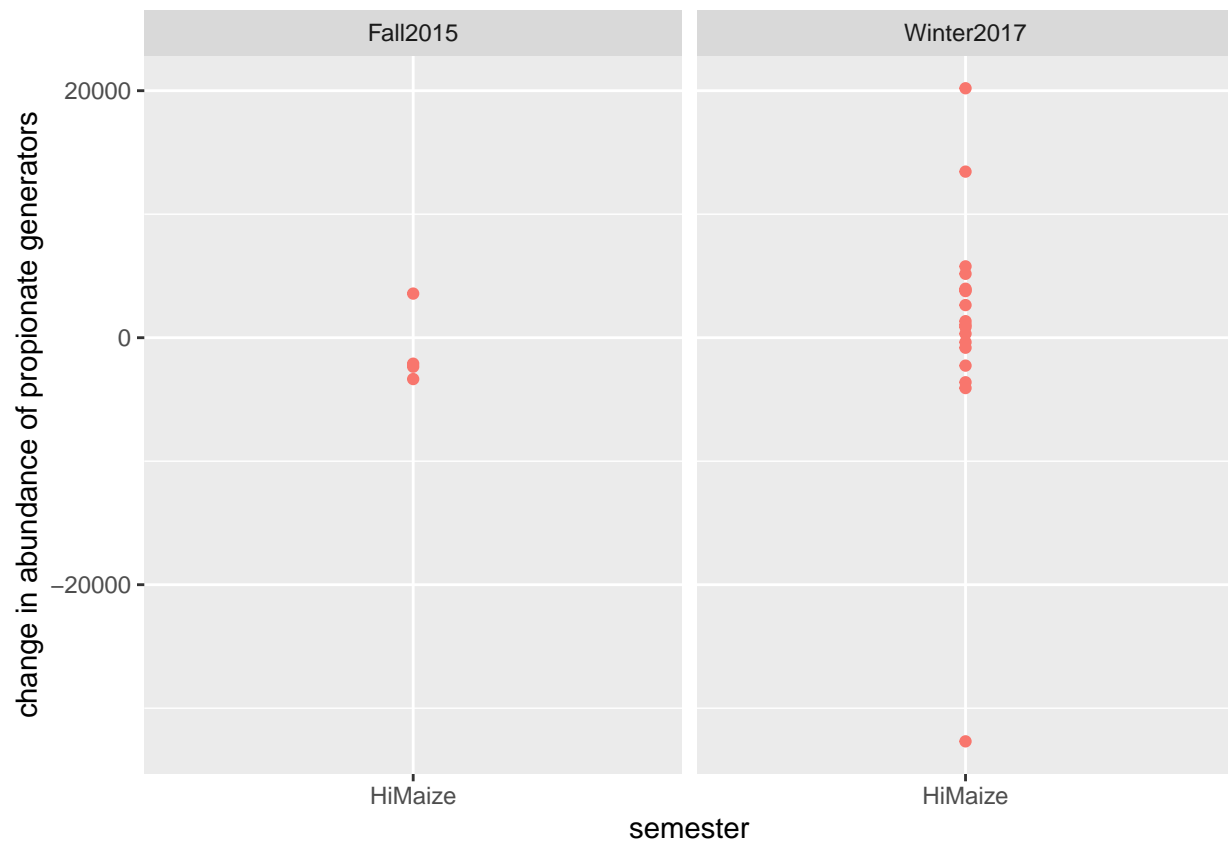
#the p value is 0.8417, larger than 0.05, which means we can't use the linear model to interpret the

Delta abundance shown by graph & Is delta abundance of two semesters students the same?

```
plot_abd <- merged_data %>%
  ggplot(aes(x = supplement_consumed,
             y = delta_abundance,
             supplement_consumed, semester)) +
  geom_point(aes(color = "delta_abundance")) +
  facet_grid(~semester, scales = "free")+
  labs(x = "semester",
       y = "change in abundance of propionate generators") +
  theme(legend.position = "none")
```

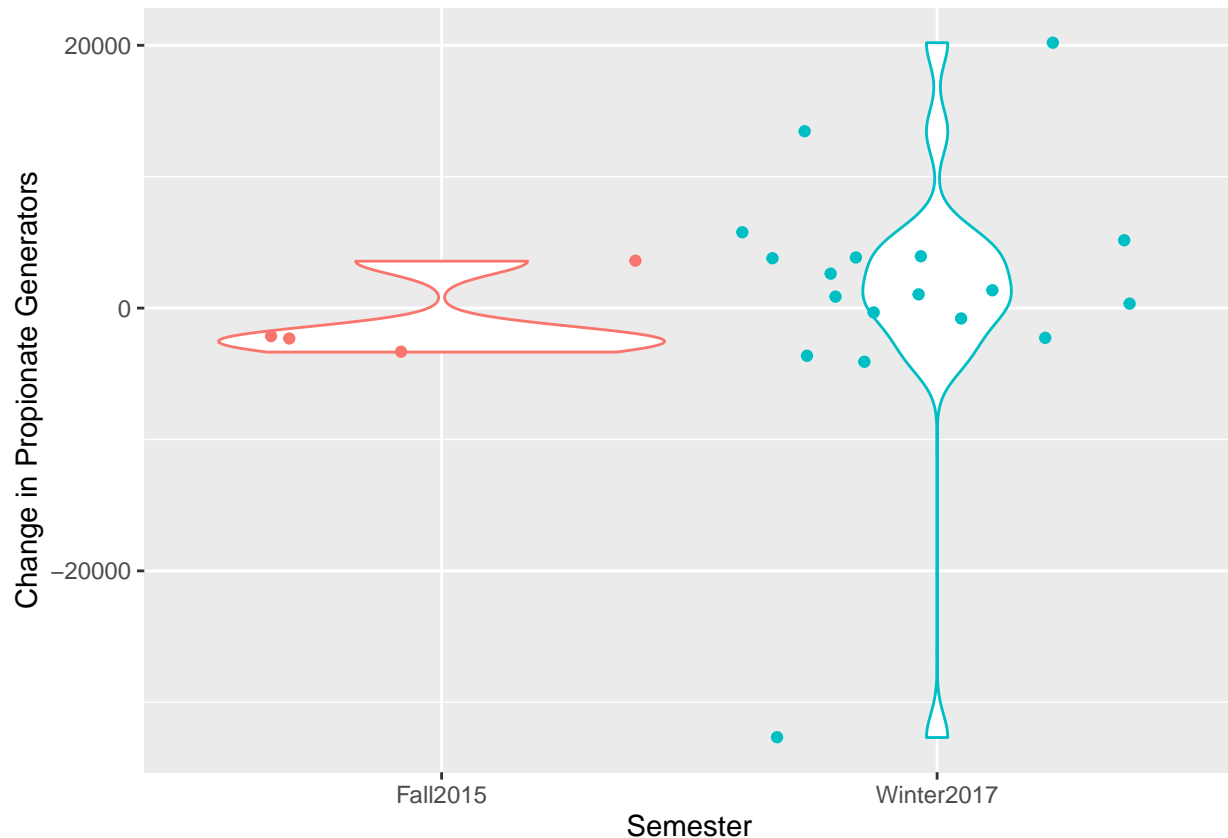
```
## Warning: Duplicated aesthetics after name standardisation:
```

```
plot_abd
```



```
sem_comparison_dabund <- merged_data %>%
  ggplot(aes(x = semester,
             y = delta_abundance,
             color = semester)) +
  geom_violin() + geom_jitter() +
  xlab("Semester") +
  ylab("Change in Propionate Generators") +
  theme(legend.position = "none")

sem_comparison_dabund
```



```
save_plot(filename = "figures/semester_delta_abund.pdf",
          plot= sem_comparison_dabund,
          nrow = 2, ncol = 3,
          base_aspect_ratio = 1.1)
save_plot(filename = "figures/delta_abundance_by_semester.pdf",
          plot= plot_abd,
          nrow = 1, ncol = 2,
          base_aspect_ratio = 1.1)
```

###Ordination plot and propionate phyloseq objects

```
# Created a phyloseq object in order to make the ordination plot.
long_priop_delta <- read_delim("~/Documents/Bio201_Project/curated_data/long_priop_delta.txt",
                               delim = "\t", escape_double = FALSE,
```

```

        trim_ws = TRUE,
        col_types = cols())

long_priop_delta_table <- long_priop_delta %>%
  mutate(sample_id = paste(participant_id, study_week, sep="_")) %>%
  column_to_rownames(var = "sample_id") %>%
  sample_data(.)

all_data_indv <- read_delim("~/Documents/Bio201_Project/raw_data/species_avg_shared.txt",
  delim = "\t", escape_double = FALSE,
  trim_ws = TRUE,
  col_types = cols()) %>%

  rename_all(tolower) %>%
  filter(semester == "Winter2017" | semester == "Fall2015") %>%
  select(-semester) %>%
  mutate(sample_id = paste(participant_id, study_week, sep="_")) %>%
  distinct(., sample_id, .keep_all = TRUE) %>%
  column_to_rownames(var = "sample_id") %>%
  select(-participant_id, -study_week) %>%
  as.matrix() %>%
  otu_table(., taxa_are_rows = FALSE)

# Merge all three data frames into phyloseq obj
physq_obj <- phyloseq(all_data_indv, long_priop_delta_table)
physq_obj

```

```

## phyloseq-class experiment-level object
## otu_table() OTU Table:          [ 1160 taxa and 49 samples ]
## sample_data() Sample Data:      [ 49 samples by 8 sample variables ]

```

```

# ordination plot
physq_bc <- ordinate(physq_obj,
  method = "NMDS",
  k=2, maxit=100, try=50,
  distance = "bray")

```

```

## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.1696927
## Run 1 stress 0.1988413
## Run 2 stress 0.1761784
## Run 3 stress 0.1733553
## Run 4 stress 0.1752794
## Run 5 stress 0.1694109
## ... New best solution
## ... Procrustes: rmse 0.04706191 max resid 0.2414224
## Run 6 stress 0.1697455
## ... Procrustes: rmse 0.04690441 max resid 0.2417186
## Run 7 stress 0.2035519
## Run 8 stress 0.1852868
## Run 9 stress 0.2026066
## Run 10 stress 0.1735293
## Run 11 stress 0.173355

```

```

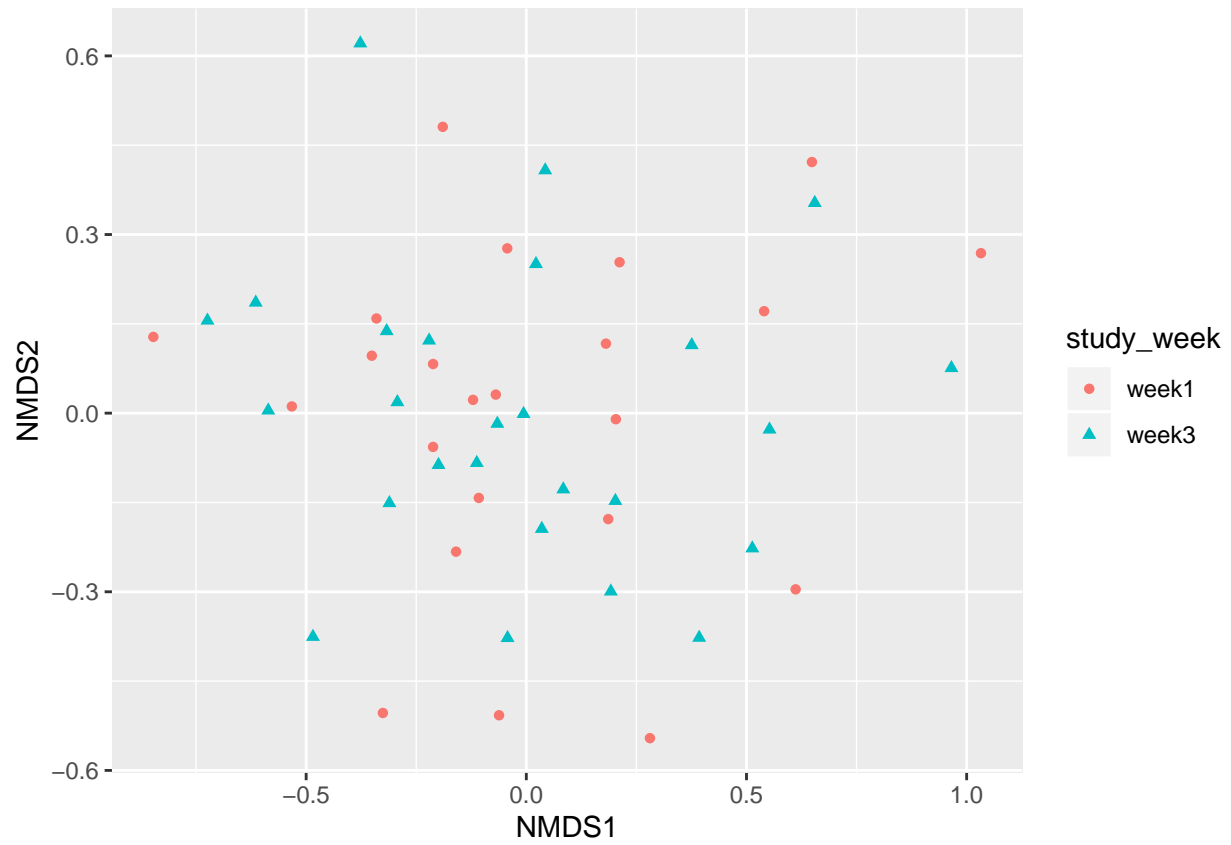
## Run 12 stress 0.1854104
## Run 13 stress 0.1860008
## Run 14 stress 0.171788
## Run 15 stress 0.2042453
## Run 16 stress 0.1697452
## ... Procrustes: rmse 0.04686196  max resid 0.2416843
## Run 17 stress 0.1901745
## Run 18 stress 0.1733555
## Run 19 stress 0.1860011
## Run 20 stress 0.1697461
## ... Procrustes: rmse 0.04679095  max resid 0.2416845
## Run 21 stress 0.1808968
## Run 22 stress 0.1855603
## Run 23 stress 0.2004145
## Run 24 stress 0.173266
## Run 25 stress 0.1754609
## Run 26 stress 0.1752792
## Run 27 stress 0.1993348
## Run 28 stress 0.1733548
## Run 29 stress 0.1709983
## Run 30 stress 0.1752792
## Run 31 stress 0.1908532
## Run 32 stress 0.1998887
## Run 33 stress 0.1949159
## Run 34 stress 0.1733563
## Run 35 stress 0.1818543
## Run 36 stress 0.1789663
## Run 37 stress 0.1752792
## Run 38 stress 0.2035656
## Run 39 stress 0.1694112
## ... Procrustes: rmse 0.0001858266  max resid 0.0009843029
## ... Similar to previous best
## Run 40 stress 0.1697453
## ... Procrustes: rmse 0.04688079  max resid 0.2416931
## Run 41 stress 0.1696935
## ... Procrustes: rmse 0.04710399  max resid 0.2410157
## Run 42 stress 0.2080632
## Run 43 stress 0.1752794
## Run 44 stress 0.1992261
## Run 45 stress 0.2087614
## Run 46 stress 0.199913
## Run 47 stress 0.1732661
## Run 48 stress 0.1919815
## Run 49 stress 0.1854285
## Run 50 stress 0.2068243
## *** Solution reached

```

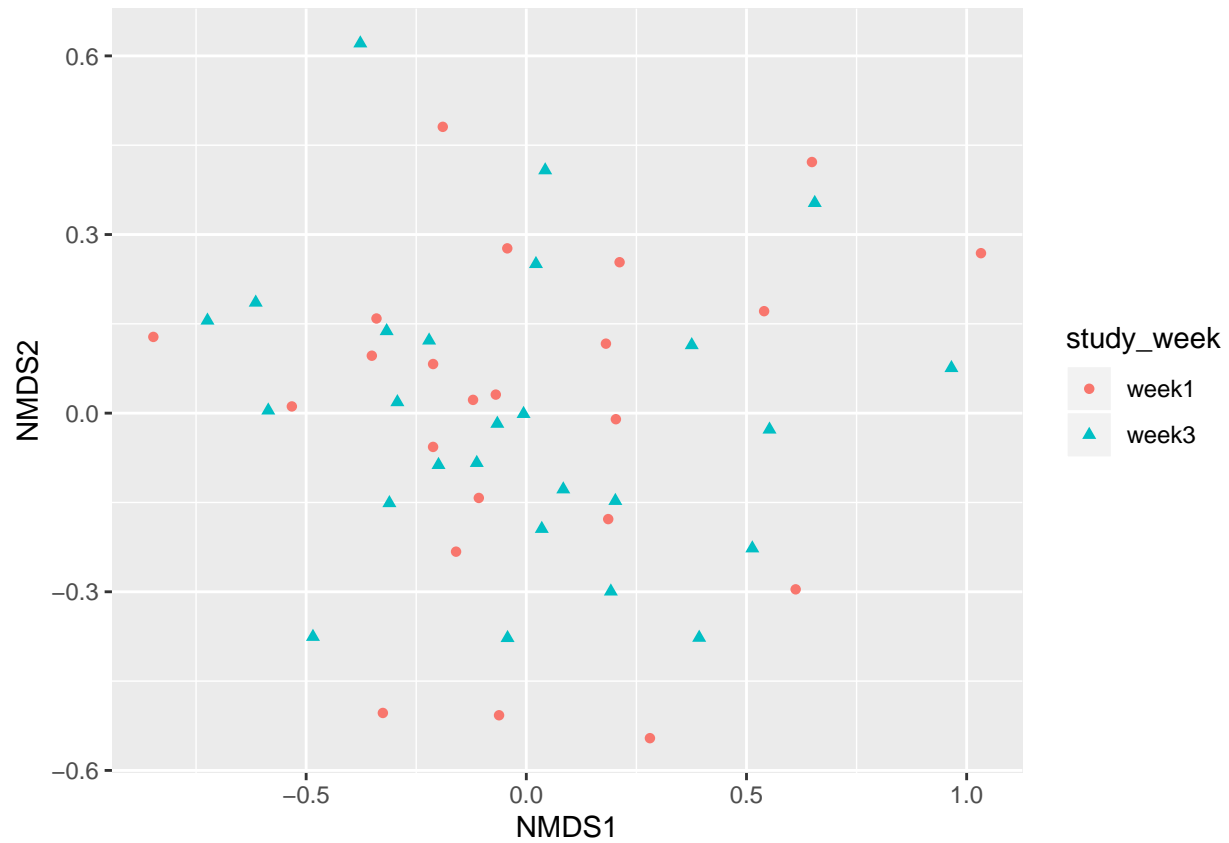
```

ordplot1 <- plot_ordination(physeq = physq_obj,
  ordination = physq_bc,
  type = "samples",
  color = "study_week",
  shape = "study_week") %>%
plot(main = "Ordination Plot of Study Week & Semester Data")

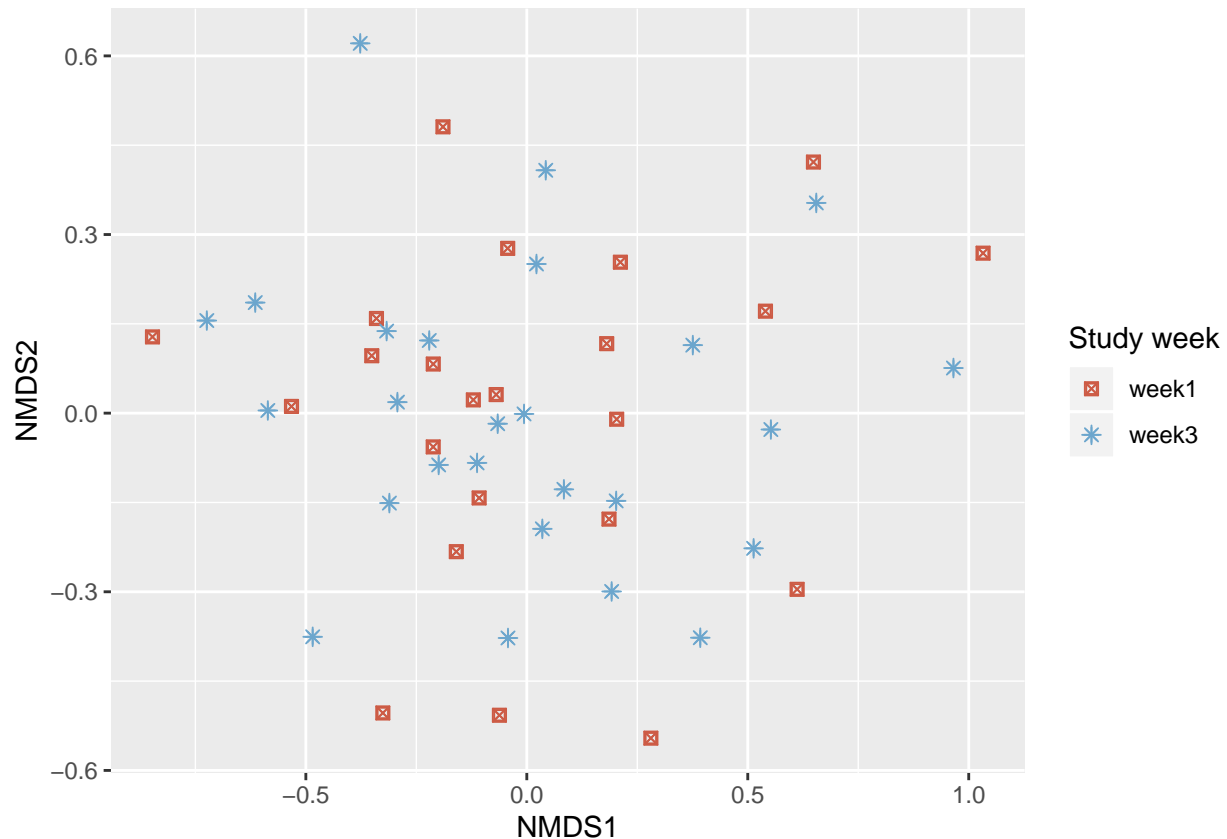
```



ordplot1



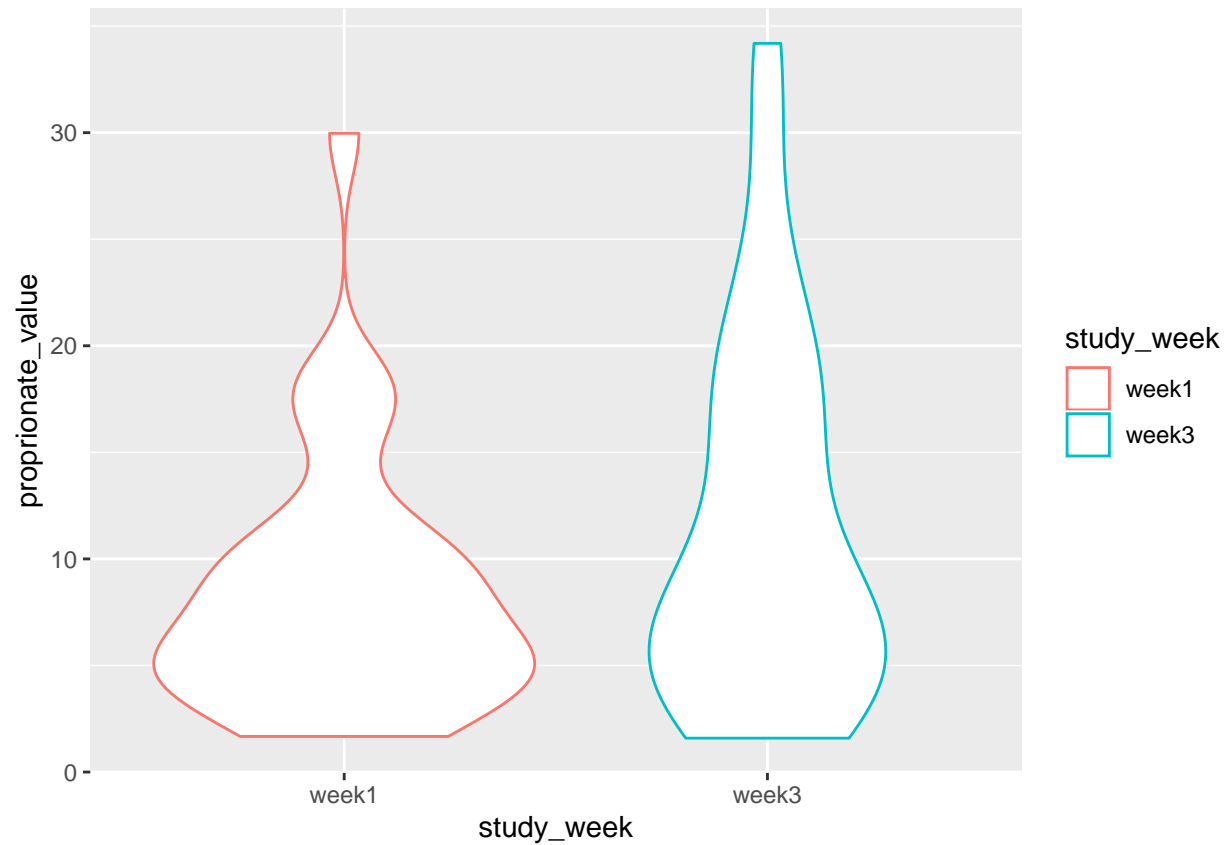
```
ordplot2 <- ordplot1 +
  geom_point(size=2) +
  scale_color_manual(name = "Study week",
    values = c("coral3",
               "skyblue3"),
    labels = c("week1", "week3")) +
  scale_shape_manual(name = "Study week",
    values = c(7, 8),
    labels = c("week1", "week3"))
ordplot2
```

```
save_plot(filename = "figures/ordplot2.pdf",
  plot = ordplot2,
  nrow = 1, ncol = 2,
  base_aspect_ratio = 1.1)
save_plot(filename = "figures/ordplot1.pdf",
  plot = ordplot1,
  nrow = 1, ncol = 2,
  base_aspect_ratio = 1.1)
```

Violin Plot -> To check whether or not the proprionate levels changed from week1 to week3, and to check how different the two semesters were. The two semesters looked very different, which could explain why our data looks very random and we found no correlation. Each semester contains very different microbiomes, all that change in different ways in response to the supplement.

```
priop_violin <- long_priop_delta %>%
  filter(study_week == "week1" | study_week == "week3") %>%
  ggplot(aes(x = study_week,
    y = proprionate_value,
    color = study_week)) +
  geom_violin()
priop_violin
```



```
priop_violin2 <- long_priop_delta %>%  
  filter(semester == "Fall2015" | semester == "Winter2017") %>%  
  ggplot(aes(x = semester,  
            y = proprionate_value,  
            color = semester)) +  
  geom_violin()  
priop_violin2
```



```
save_plot(filename = "figures/priop_violin.pdf",
  plot = priop_violin,
  nrow = 1, ncol = 2,
  base_aspect_ratio = 1.1)
save_plot(filename = "figures/priop_violin2.pdf",
  plot = priop_violin2,
  nrow = 1, ncol = 2,
  base_aspect_ratio = 1.1)
```

#Bar plot to show the different bacterial compositions of the two semesters (phyloseq objects to bar pl

```
phylum_avg_shared <- read_delim("~/Documents/Bio201_Project/raw_data/phylum_avg_shared.txt",
  delim = "\t", escape_double = FALSE,
  trim_ws = TRUE,
  col_types = cols())
```

```
phylum_data <- phylum_avg_shared %>%
  rename_all(tolower) %>%
  filter(study_week == "week1" | study_week == "week3",
    semester == "Fall2015" | semester == "Winter2017") %>%
  mutate(sample_id = paste(participant_id, study_week, sep = "_")) %>%
  select(-participant_id, -study_week, -semester) %>%
  column_to_rownames(var = "sample_id") %>%
  as.matrix() %>% #
  otu_table(., taxa_are_rows = FALSE)
```

```

scfa_weekly_wide <- read_delim(file = "~/Documents/Bio201_Project/final_project_data/DB_v_0.08/SCFA_wk1",
                             delim = "\t", col_names = TRUE, trim_ws = TRUE,
                             na = c("", "NA")) %>%
  rename_all(tolower) %>%
  filter(quantity_compliant == "yes", study_week == "week1" | study_week == "week3") %>%
  select(-starts_with("ace"), -starts_with("but"), -ends_with("median"))

```

```

## Parsed with column specification:
## cols(
##   Participant_ID = col_character(),
##   Study_week = col_character(),
##   Frequency = col_character(),
##   Semester = col_character(),
##   Supplement_consumed = col_character(),
##   Quantity_compliant = col_character(),
##   Acetate_median = col_double(),
##   Acetate_mean = col_double(),
##   Butyrate_median = col_double(),
##   Butyrate_mean = col_double(),
##   Propionate_median = col_double(),
##   Propionate_mean = col_double()
## )

```

```

HiMaize_weekly_wide <- scfa_weekly_wide %>%
  filter(!is.na(supplement_consumed)) %>%
  filter(supplement_consumed == "HiMaize") %>%
  mutate(sample_id = paste(participant_id, study_week, sep = "_")) %>%
  column_to_rownames(var = "sample_id") %>%
  sample_data(.)

```

```

physq_phylum_obj <- phyloseq(phylum_data, HiMaize_weekly_wide)
physq_phylum_obj

```

```

## phyloseq-class experiment-level object
## otu_table() OTU Table:          [ 35 taxa and 65 samples ]
## sample_data() Sample Data:      [ 65 samples by 7 sample variables ]

```

```

physq_long <- physq_phylum_obj %>%
  transform_sample_counts(function(x) {x/sum(x)} ) %>%
  psmelt() %>%
  rename_all(tolower)

```

```

phylum_plot <- ggplot(data = physq_long,
  aes(x = study_week,
    y = abundance,
    fill = otu)) +
  geom_col() +
  facet_grid(~semester) +
  xlab(NULL) +
  theme(axis.text.x = element_text(angle = 90,
    vjust = 0.5)) +
  ylab("Relative Abundance (Phyla > 0.1%)")
phylum_plot

```



```
save_plot(filename = "figures/phylum_plot.pdf",
          plot = phylum_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 3)
```

Conclusion

From our data, we observed that only one supplement, HiMaize, produced a possibly significant change in propionate concentration from week 1 to week 3. Further analysis and testing of the HiMaize data offered no other significant results or correlations, but this may be in large part due to the inconsistent and variable nature of the data collected. Further, substantial research on this topic will require, most importantly, a much larger sample size. We believe propionate increase is a productive topic to continue research in because of its positive impacts on diet and human health.