Group1_finalprojectRmd

Set Working Directory

Load Packages

```r
library(tidyverse)
library(readxl)
library(broom)
library(cowplot)
library(phyloseq)
library(vegan)
set.seed(7)
```

#Import Data

```r
library(readr)
biographical <- read_delim("~/Group1_Bio201_FinalProject/raw_data/biographical.txt",
    "\t", escape_double = FALSE, trim_ws = TRUE)
```

```
## Parsed with column specification:
## cols(
##   Participant_ID = col_character(),
##   Use_Data = col_character(),
##   Sex = col_character(),
##   Age = col_double(),
##   Race_ethnicity = col_character(),
##   Weight_kg = col_double(),
##   Height_meters = col_double(),
##   BMI = col_double()
## )
```

```r
View(biographical)

library(readr)
phylum_avg_shared <- read_delim("~/Group1_Bio201_FinalProject/raw_data/phylum_avg_shared.txt",
    "\t", escape_double = FALSE, trim_ws = TRUE)
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Participant_ID = col_character(),
##   Semester = col_character(),
##   Study_week = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```r
View(phylum_avg_shared)

library(readr)
all_data_wkly <- read_delim("~/Group1_Bio201_FinalProject/raw_data/Bigzipfile/all_data_wkly.txt",
    "\t", escape_double = FALSE, trim_ws = TRUE)
```

```
## Parsed with column specification:
## cols(
##   Participant_ID = col_character(),
##   Study_week = col_character(),
##   Semester = col_character(),
##   Supplement_consumed = col_character(),
##   Quantity_compliant = col_character(),
##   Frequency = col_character(),
##   pH_median = col_double(),
##   pH_mean = col_double(),
##   Bristol_median = col_logical(),
##   Bristol_mean = col_logical(),
##   Blood_glucose_median = col_double(),
##   Blood_glucose_mean = col_double(),
##   Acetate_median = col_double(),
##   Acetate_mean = col_double(),
##   Butyrate_median = col_double(),
##   Butyrate_mean = col_double(),
##   Propionate_median = col_double(),
##   Propionate_mean = col_double()
## )
```

```
## Warning: 755 parsing failures.
##  row           col               expected actual
## 1066 Bristol_median 1/0/T/F/TRUE/FALSE      4 '~/Group1_Bio201_FinalProject/raw_data/Bigzipfile/all_
## 1066 Bristol_mean   1/0/T/F/TRUE/FALSE      4 '~/Group1_Bio201_FinalProject/raw_data/Bigzipfile/all_
## 1067 Bristol_median 1/0/T/F/TRUE/FALSE      5 '~/Group1_Bio201_FinalProject/raw_data/Bigzipfile/all_
## 1067 Bristol_mean   1/0/T/F/TRUE/FALSE      5 '~/Group1_Bio201_FinalProject/raw_data/Bigzipfile/all_
## 1068 Bristol_median 1/0/T/F/TRUE/FALSE      5 '~/Group1_Bio201_FinalProject/raw_data/Bigzipfile/all_
## .... .............. .................. ...... .....................................................
## See problems(...) for more details.
```

```r
View(all_data_wkly)
```

#Tidy Data ##Butyrate and Propionate df

```r
c_all_data_wkly <- all_data_wkly %>%
  rename_all(tolower) %>%
  select(participant_id, semester, supplement_consumed, frequency, quantity_compliant, study_week, buty
  filter(study_week == "week1" | study_week == "week3", quantity_compliant == "yes") #curated all data

#wide dfs
but_wide <- all_data_wkly %>%
  rename_all(tolower) %>%
  select(participant_id, semester, supplement_consumed, frequency, quantity_compliant, study_week, buty
  filter(study_week == "week1" | study_week == "week3", quantity_compliant == "yes") %>% #keep only wee
```

```
    spread(study_week, butyrate_mean) %>%
    mutate(delta_butyrate = week3 - week1) #calculate difference in butyrate

pro_wide <- all_data_wkly %>%
  rename_all(tolower) %>%
  select(participant_id, semester, supplement_consumed, frequency, quantity_compliant, study_week, prop
  filter(study_week == "week1" | study_week == "week3", quantity_compliant == "yes") %>% #keep only wee
  spread(study_week, propionate_mean) %>%
  mutate(delta_propionate = week3 - week1) #calculate difference in propionate

but_pro_wide <- inner_join(x = but_wide, y = pro_wide,
                  by = c("participant_id",
                         "semester", "supplement_consumed", "frequency")) %>%
  rename(but_wk1 = week1.x, but_wk3 = week3.x, pro_wk1 = week1.y, pro_wk3 = week3.y) %>%
  select(-starts_with("quantity_compliant")) #all measurements by participant ID

#Save Data Frames:
write_delim(c_all_data_wkly, path = "curated_data/c_all_data_wkly.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(but_wide, path = "curated_data/but_wide.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(pro_wide, path = "curated_data/pro_wide.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(but_pro_wide, path = "curated_data/but_pro_wide.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
```

BMI df

```
BMI_data <- biographical %>%
  rename_all(tolower) %>%
  select(participant_id, sex, age, race_ethnicity, weight_kg, height_meters) %>%
  drop_na(weight_kg, height_meters) %>%
  mutate(BMI = weight_kg/(height_meters)^2) #calculate all BMIs; n = 252

underweight_BMI <- BMI_data %>%
  filter(BMI < 18.5) #subset by CDC category: underweight - n = 12

healthy_BMI <- BMI_data %>%
  filter(BMI >= 18.5, BMI < 25) #subset by CDC category: healthy - n = 186

overweight_BMI <- BMI_data %>%
  filter(BMI >= 25, BMI < 30) #subset by CDC category: overweight - n = 39

obese_BMI <- BMI_data %>%
  filter(BMI >= 30) #subset by CDC category: obese - n = 15

#total n = 252

#Save Data Frames
write_delim(BMI_data, path = "curated_data/BMI_data.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(underweight_BMI, path = "curated_data/underweight_BMI.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
```

```
write_delim(healthy_BMI, path = "curated_data/healthy_BMI.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(overweight_BMI, path = "curated_data/overweight_BMI.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(obese_BMI, path = "curated_data/obese_BMI.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
```

Phylum df

```
#Creating Processed Phylum Data Frame:
phylum_processed_df <- phylum_avg_shared %>%
  rename_all(tolower) %>%
  filter(study_week == "week1" | study_week == "week3") %>%
  select("participant_id", "semester", "study_week", "firmicutes", "bacteroidetes <bacteroidetes>") %>%
  rename(., "bacteroidetes" = "bacteroidetes <bacteroidetes>") %>%
  mutate(., f_b_ratio = firmicutes/bacteroidetes)

#Checking Sample Sizes for Consistency by Week:
phylum_processed_df %>%
  filter(study_week == "week1") %>%
  summarise(sample_size = n()) # n = 508
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1         508
```

```
phylum_processed_df %>%
  filter(study_week == "week3") %>%
  summarise(sample_size = n()) # n = 480. 508-480 = 28 inconsistencies.
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1         480
```

```
#Split by Weeks and Recombine so only samples with both weeks present:
phylum_week1 <- phylum_processed_df %>%
  filter(study_week == "week1")

phylum_week3 <- phylum_processed_df %>%
  filter(study_week == "week3")
phylum_week3
```

```
## # A tibble: 480 x 6
##    participant_id semester study_week firmicutes bacteroidetes f_b_ratio
##    <chr>          <chr>    <chr>           <dbl>         <dbl>     <dbl>
## 1 U042           Fall2015 week3           13515         16756     0.807
## 2 U044           Fall2015 week3           15632          7587     2.06
## 3 U046           Fall2015 week3            7980.        17707     0.451
## 4 U048           Fall2015 week3           12427         22233     0.559
```

```
##  5 U050      Fall2015 week3        12079         9706      1.24
##  6 U052      Fall2015 week3         5324        16858      0.316
##  7 U054      Fall2015 week3        12668.       14542.     0.871
##  8 U057      Fall2015 week3         8265        16255      0.508
##  9 U059      Fall2015 week3         8742        14673      0.596
## 10 U062      Fall2015 week3         6117.       21792.     0.281
## # ... with 470 more rows
```

```r
phylum_processed_df_both_weeks_present <- inner_join(phylum_week1, phylum_week3, by = (c("participant_i
```

```r
#Final Data Frames:
phylum_processed_df
```

```
## # A tibble: 988 x 6
##    participant_id semester study_week firmicutes bacteroidetes f_b_ratio
##    <chr>          <chr>    <chr>           <dbl>         <dbl>     <dbl>
##  1 U042           Fall2015 week1           14119         18149     0.778
##  2 U042           Fall2015 week3           13515         16756     0.807
##  3 U044           Fall2015 week3           15632          7587     2.06
##  4 U046           Fall2015 week3            7980.        17707     0.451
##  5 U048           Fall2015 week3           12427         22233     0.559
##  6 U050           Fall2015 week3           12079          9706     1.24
##  7 U051           Fall2015 week1           19352          4827     4.01
##  8 U052           Fall2015 week1            4944         13166     0.376
##  9 U052           Fall2015 week3            5324         16858     0.316
## 10 U053           Fall2015 week1            7061         22655     0.312
## # ... with 978 more rows
```

```r
phylum_processed_df_both_weeks_present
```

```
## # A tibble: 451 x 11
##    participant_id semester.x study_week.x firmicutes.x bacteroidetes.x
##    <chr>          <chr>      <chr>              <dbl>           <dbl>
##  1 U042           Fall2015   week1              14119           18149
##  2 U052           Fall2015   week1               4944           13166
##  3 U054           Fall2015   week1              11479           20357
##  4 U062           Fall2015   week1               8366           25134.
##  5 U063           Fall2015   week1              16913.          10266
##  6 U064           Fall2015   week1              10583           10990
##  7 U066           Fall2015   week1               9209           14273
##  8 U067           Fall2015   week1              19163.          16210.
##  9 U068           Fall2015   week1               6034           20364
## 10 U070           Fall2015   week1              11963           13446
## # ... with 441 more rows, and 6 more variables: f_b_ratio.x <dbl>,
## #   semester.y <chr>, study_week.y <chr>, firmicutes.y <dbl>,
## #   bacteroidetes.y <dbl>, f_b_ratio.y <dbl>
```

```r
phylum_processed_df_both_weeks_present_final <- phylum_processed_df_both_weeks_present %>%
  rename(firmicutes_wk1 = firmicutes.x, bacteroidetes_wk1 = bacteroidetes.x, fb_ratio_wk1 =
  f_b_ratio.x, firmicutes_wk3 = firmicutes.y, bacteroidetes_wk3 = bacteroidetes.y, fb_ratio_wk3 =
  f_b_ratio.y, semester = semester.x) %>%
  select(-starts_with("study_week"), -starts_with("semester.y"))
```

```r
View(phylum_processed_df_both_weeks_present_final)

#Save data frames:
write_delim(phylum_processed_df, path = "curated_data/phylum_processed_df.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(phylum_processed_df_both_weeks_present, path = "curated_data/phylum_processed_df_both_weeks_
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(phylum_processed_df_both_weeks_present_final, path = "curated_data/phylum_processed_df_both_
            delim = "\t", col_names = TRUE, quote = FALSE)
```

Joined BMI and scfas

```r
BMI_but_pro <- inner_join(x = BMI_data, y = but_pro_wide,
                          by = c("participant_id")) #n = 178

underweight_but_pro <- inner_join(x= underweight_BMI, y = but_pro_wide,
                                  by = c("participant_id")) #n = 10

healthy_but_pro <- inner_join(x= healthy_BMI, y = but_pro_wide,
                              by = c("participant_id")) #n = 133

overweight_but_pro <-  inner_join(x= overweight_BMI, y = but_pro_wide,
                                  by = c("participant_id")) #n = 25

obese_but_pro <-  inner_join(x= obese_BMI, y = but_pro_wide,
                             by = c("participant_id")) #n = 10
#total n = 178

#Save Data Frames
write_delim(BMI_but_pro, path = "curated_data/BMI_but_pro.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(underweight_but_pro, path = "curated_data/underweight_but_pro.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(healthy_but_pro, path = "curated_data/healthy_but_pro.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(overweight_but_pro, path = "curated_data/overweight_but_pro.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(obese_but_pro, path = "curated_data/obese_but_pro.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
```

Joined BMI, scfas, and phylums

```r
BMI_phylum_scfa <- inner_join(x = BMI_but_pro, y = phylum_processed_df_both_weeks_present_final,
                          by = c("participant_id")) #n = 139

underweight_phylum_scfa  <- inner_join(x = underweight_but_pro, y = phylum_processed_df_both_weeks_prese
                          by = c("participant_id")) #n = 9

healthy_phylum_scfa <- inner_join(x = healthy_but_pro, y = phylum_processed_df_both_weeks_present_final
                          by = c("participant_id")) #n = 105

overweight_phylum_scfa <- inner_join(x = overweight_but_pro, y = phylum_processed_df_both_weeks_present
                          by = c("participant_id")) #n = 20
```

```
obese_phylum_scfa <- inner_join(x = obese_but_pro, y = phylum_processed_df_both_weeks_present_final,
                    by = c("participant_id")) #n = 5

#Save Data Frames
write_delim(BMI_phylum_scfa, path = "curated_data/BMI_phylum_scfa.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(underweight_phylum_scfa, path = "curated_data/underweight_phylum_scfa.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(healthy_phylum_scfa, path = "curated_data/healthy_phylum_scfa.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(overweight_phylum_scfa, path = "curated_data/overweight_phylum_scfa.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
write_delim(obese_phylum_scfa, path = "curated_data/obese_phylum_scfa.txt",
            delim = "\t", col_names = TRUE, quote = FALSE)
```

#Analysis ##Broad (All BMIs) ###F:B and BMI

```
lm_Broad_BMI_plot_wk1<-BMI_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = BMI)) +
  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("FB_Ratio") +
  ylab("BMI") + labs(title = "Broad BMI and F:B Ratio Correlation wk1")
print(lm_Broad_BMI_plot_wk1)
```

Broad BMI and F:B Ratio Correlation wk1
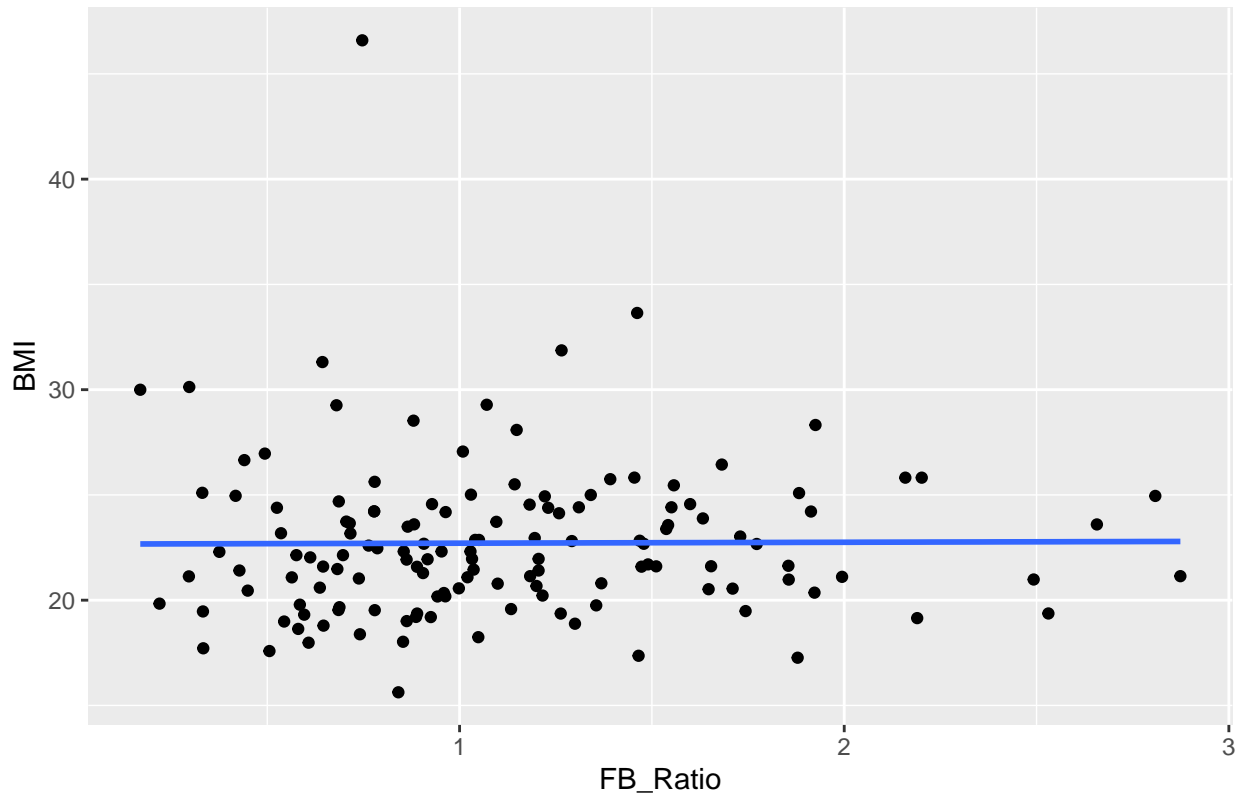

```
lm_Broad_BMI_wk1<-BMI_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ BMI, data = .) #test relationship
summary(lm_Broad_BMI_wk1)
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ BMI, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95021 -0.40410 -0.09109  0.34665  1.76222
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0918175  0.2875401   3.797 0.000219 ***
## BMI         0.0009424  0.0124923   0.075 0.939977
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5505 on 137 degrees of freedom
## Multiple R-squared:  4.154e-05,  Adjusted R-squared:  -0.007257
## F-statistic: 0.005691 on 1 and 137 DF,  p-value: 0.94
```

```
lm_Broad_BMI_plot_wk3<-BMI_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
```

```
            y = BMI)) +
  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("FB_Ratio") +
  ylab("BMI")  + labs(title = "Broad BMI and F:B Ratio Correlation wk3")
print(lm_Broad_BMI_plot_wk3)
```

Broad BMI and F:B Ratio Correlation wk3
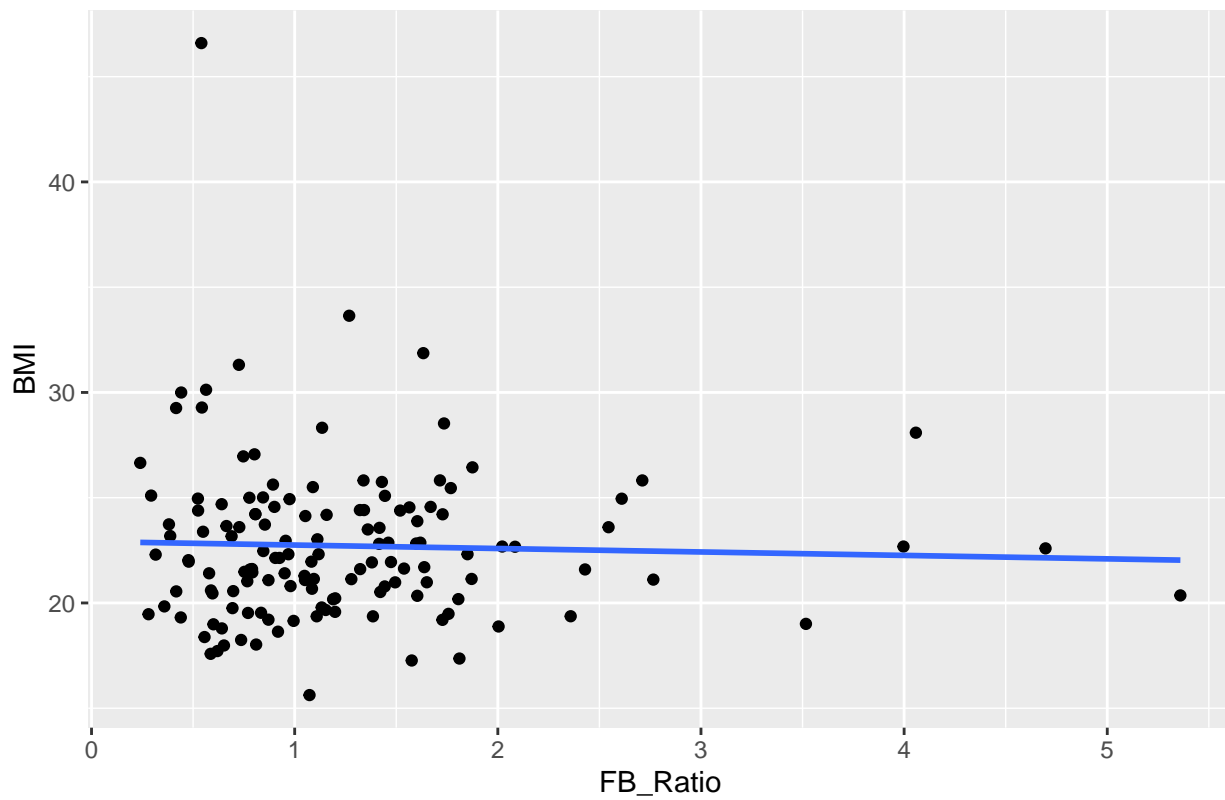

```
lm_Broad_wk3_BMI<-BMI_phylum_scfa %>%
  lm(fb_ratio_wk3 ~ BMI, data = .) #test relationship
summary(lm_Broad_wk3_BMI)
```

```
##
## Call:
## lm(formula = fb_ratio_wk3 ~ BMI, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9814 -0.4982 -0.1734  0.2963  4.1048
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.412211   0.426001   3.315  0.00117 **
## BMI         -0.007715   0.018508  -0.417  0.67745
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8157 on 137 degrees of freedom
## Multiple R-squared:  0.001267,   Adjusted R-squared:  -0.006023
## F-statistic: 0.1737 on 1 and 137 DF,  p-value: 0.6775
```

```r
save_plot(filename = "figures/lm_Broad_BMI_plot_wk1.pdf",
          plot = lm_Broad_BMI_plot_wk1,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)

save_plot(filename = "figures/lm_Broad_BMI_plot_wk3.pdf",
          plot = lm_Broad_BMI_plot_wk3,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the BMI of all the participants is .94, which is not statistically significant. The R^2 value was also insignificant being very small at $R^2 = -.0073$. There is no significant correlation between the week 1 firmicutes - bacteroidetes ratio and BMI.

Similarly, the week 3 F - B ratio and BMI also had an insignifcant p-value and R^2 value at $p=.67$ and $R^2 = -.006$ respectively. This also represents no significant correlation.

###F:B and Butyrate

```r
lm_wk1_but_Broad_plot<-BMI_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = but_wk1)) +
  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("F:B ratio wk1") +
  ylab("Butyrate Concentration kg/mmol wk1")  + labs(title = "Broad F:B Ratio and Butyrate Concentration
print(lm_wk1_but_Broad_plot)
```

Broad F:B Ratio and Butyrate Concentration Correlation wk1
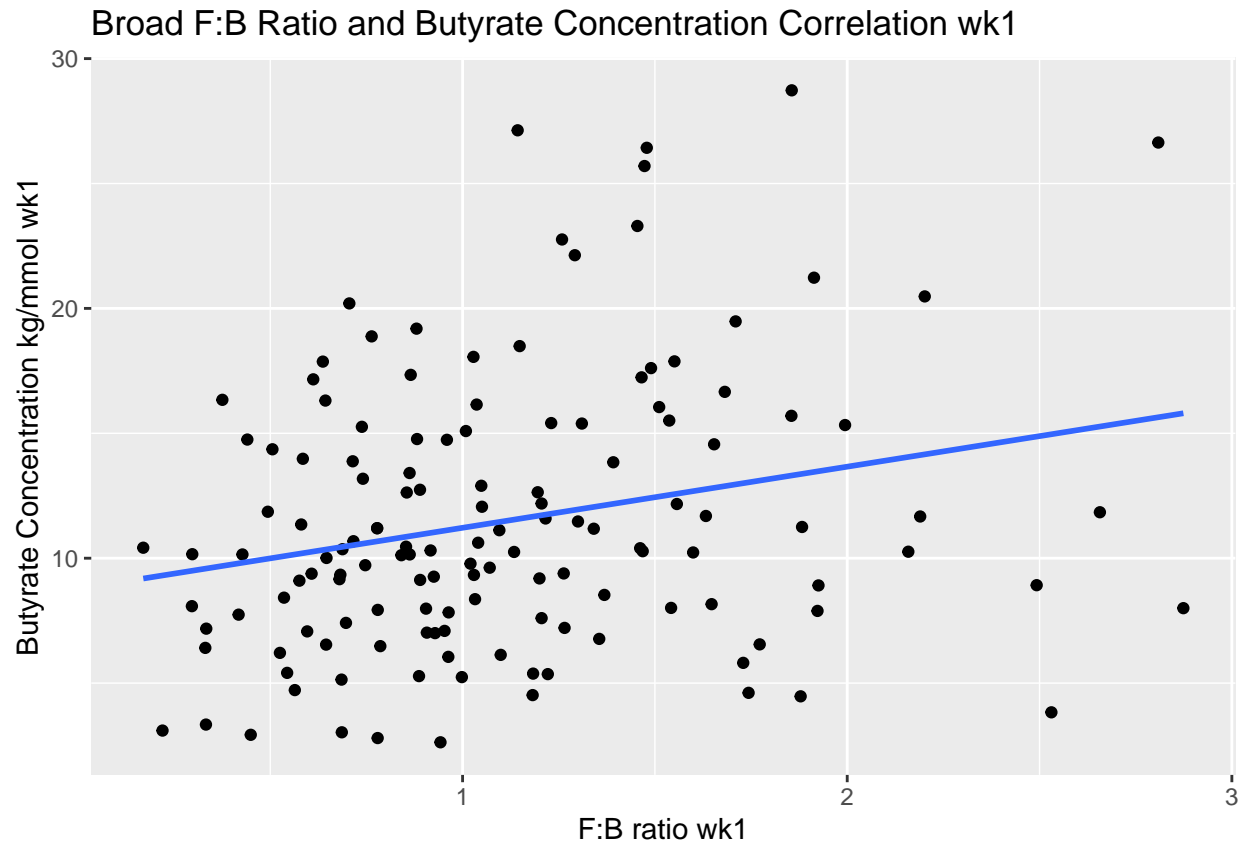

```
lm_wk1_but_Broad<-BMI_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ but_wk1, data = .) #test relationship
summary(lm_wk1_but_Broad)
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ but_wk1, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.91708 -0.38098 -0.07662  0.25754  1.84617
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.832191   0.105412   7.895 8.39e-13 ***
## but_wk1     0.024450   0.008282   2.952  0.00371 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5338 on 137 degrees of freedom
## Multiple R-squared:  0.05981,    Adjusted R-squared:  0.05295
## F-statistic: 8.716 on 1 and 137 DF,  p-value: 0.003713
```

```
lm_wk3_but_Broad_plot<-BMI_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
```

```
              y = but_wk3)) +
  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("F:B ratio wk3") +
  ylab("Butyrate Concentration kg/mmol wk3")  + labs(title = "Broad F:B Ratio and Butyrate Concentration
print(lm_wk3_but_Broad_plot)
```

Broad F:B Ratio and Butyrate Concentration Correlation wk3
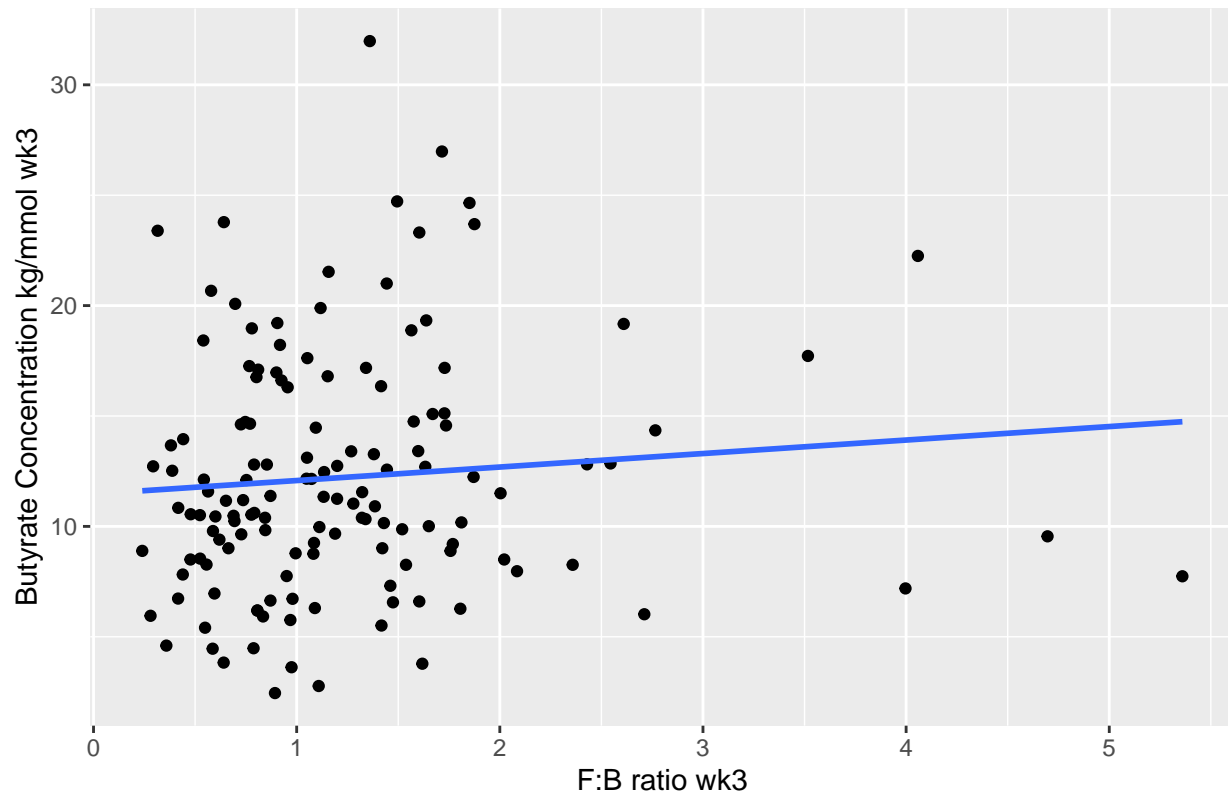

```
lm_wk3_but_Broad<-BMI_phylum_scfa %>%
  lm(fb_ratio_wk3 ~ but_wk3, data = .) #test relationship
summary(lm_wk3_but_Broad)
```

```
##
## Call:
## lm(formula = fb_ratio_wk3 ~ but_wk3, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0730 -0.5200 -0.1801  0.2977  4.1839
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.07089    0.16967   6.311 3.58e-09 ***
## but_wk3      0.01359    0.01269   1.071    0.286
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8128 on 137 degrees of freedom
## Multiple R-squared:  0.008309,   Adjusted R-squared:  0.00107
## F-statistic: 1.148 on 1 and 137 DF,  p-value: 0.2859
```

```r
save_plot(filename = "figures/lm_wk1_but_Broad_plot.pdf",
          plot = lm_wk1_but_Broad_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.2)

save_plot(filename = "figures/lm_wk3_but_Broad_plot.pdf",
          plot = lm_wk3_but_Broad_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.2)
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the week 1 Butyrate concentration of all the participants is .0037, which is statistically significant. However, the R^2 value is insignificant being very small at R^2=0.0053, which represents too much variation in the data. There is no significant correlation between the week 1 firmicutes - bacteroidetes ratio and Butyrate Concentration.

The week 3 F - B ratio and week 3 Butyrate concentration of all the participants has a insignifcant p-value and R^2 value at p=.29 and R^2 = 0.0011 respectively. This also represents no significant correlation.

###F:B and Propionate

```r
lm_wk1_pro_Broad_plot<-BMI_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = pro_wk1)) +
  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("F:B ratio wk1") +
  ylab("Propionate Concentration kg/mmol wk1")  + labs(title = "Broad F:B Ratio and Propionate Concentra
print(lm_wk1_pro_Broad_plot)
```

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

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

Broad F:B Ratio and Propionate Concentration Correlation wk1
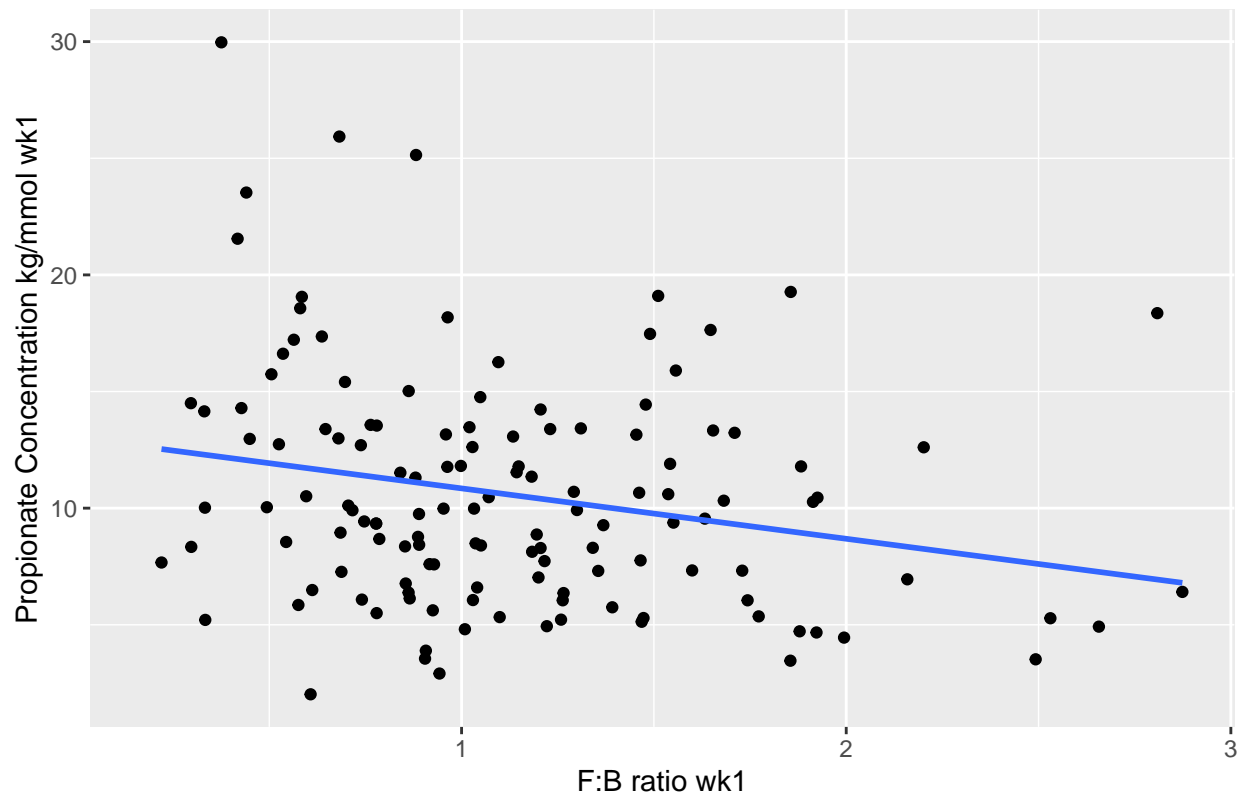

```
lm_wk1_pro_Broad<-BMI_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ pro_wk1, data = .) #test relationship
summary(lm_wk1_pro_Broad)
```
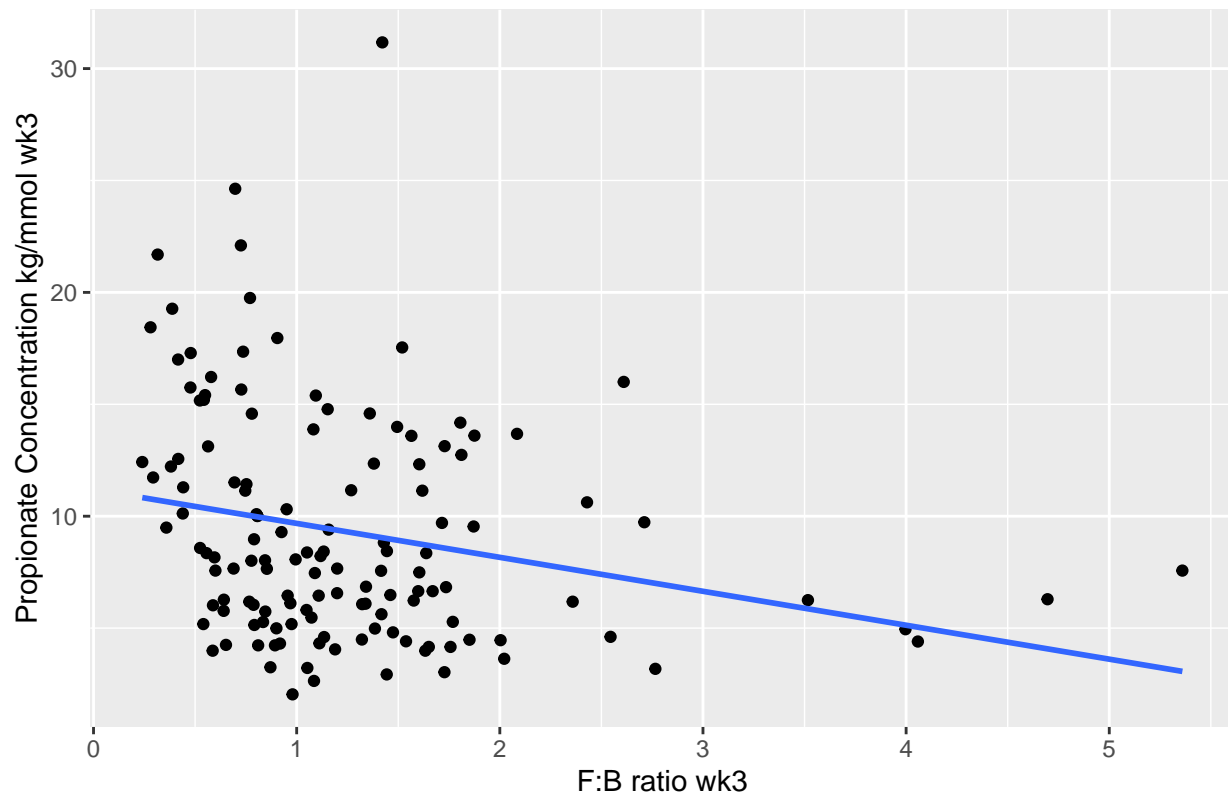
```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ pro_wk1, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97952 -0.37792 -0.09592  0.26827  1.88128
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.394523   0.107700   12.95  < 2e-16 ***
## pro_wk1     -0.025437   0.009216   -2.76  0.00661 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5282 on 131 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.05496,    Adjusted R-squared:  0.04774
## F-statistic: 7.618 on 1 and 131 DF,  p-value: 0.006607
```

```
lm_wk3_pro_Broad_plot<-BMI_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
             y = pro_wk3))+
  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("F:B ratio wk3") +
  ylab("Propionate Concentration kg/mmol wk3")  + labs(title = "Broad F:B Ratio and Propionate Concentra
print(lm_wk3_pro_Broad_plot)
```

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

```
## Warning: Removed 4 rows containing missing values (geom_point).
```



```
lm_wk3_pro_Broad<-BMI_phylum_scfa %>%
  lm(fb_ratio_wk3 ~ pro_wk3, data = .) #test relationship
summary(lm_wk3_pro_Broad)
```

```
##
## Call:
## lm(formula = fb_ratio_wk3 ~ pro_wk3, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -0.8853 -0.5050 -0.2312  0.2681  4.0435
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.61489    0.14364  11.243  < 2e-16 ***
## pro_wk3     -0.03942    0.01356  -2.908  0.00426 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7994 on 133 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.05978,    Adjusted R-squared:  0.05271
## F-statistic: 8.457 on 1 and 133 DF,  p-value: 0.004263
```

```
save_plot(filename = "figures/lm_wk1_pro_Broad_plot.pdf",
          plot = lm_wk1_pro_Broad_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

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

```
## Warning: Removed 6 rows containing missing values (geom_point).
```

```
save_plot(filename = "figures/lm_wk3_pro_Broad_plot.pdf",
          plot = lm_wk3_pro_Broad_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

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

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the week 1 Propionate concentration of all the participants is .0066, which is statistically significant. However, the R^2 value is insignificant being very small at R^2=0.0048, which represents too much variation in the data. There is no significant correlation between the week 1 firmicutes - bacteroidetes ratio and Propionate concentration.

Similarly, The week 3 F - B ratio and week 3 Propionate concentration of all the participants has a significant p-value and insignificant R^2 value at p=.0043 and R^2 = .053 respectively. This also represents no significant correlation.

##Underweight ###F:B and BMI

```
lm_under_BMI_wk1<-underweight_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ BMI, data = .) #test relationship
summary(lm_under_BMI_wk1)
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ BMI, data = .)
##
## Residuals:
```

```
##       Min       1Q   Median       3Q      Max
## -0.57151 -0.27914 -0.09574  0.19793  0.92787
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.7263     3.8631   0.706    0.503
## BMI          -0.1028     0.2196  -0.468    0.654
##
## Residual standard error: 0.5116 on 7 degrees of freedom
## Multiple R-squared:  0.03036,    Adjusted R-squared:  -0.1082
## F-statistic: 0.2192 on 1 and 7 DF,  p-value: 0.6539
```

```r
#p-value = 0.6539
#R-squared = -0.1082

lm_under_BMI_plot_wk1<-underweight_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = BMI)) +
  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("FB_Ratio") +
  ylab("BMI")  + labs(title = "Underweight BMI and F:B Ratio Correlation wk1")
print(lm_under_BMI_plot_wk1)
```


Underweight BMI and F:B Ratio Correlation wk1
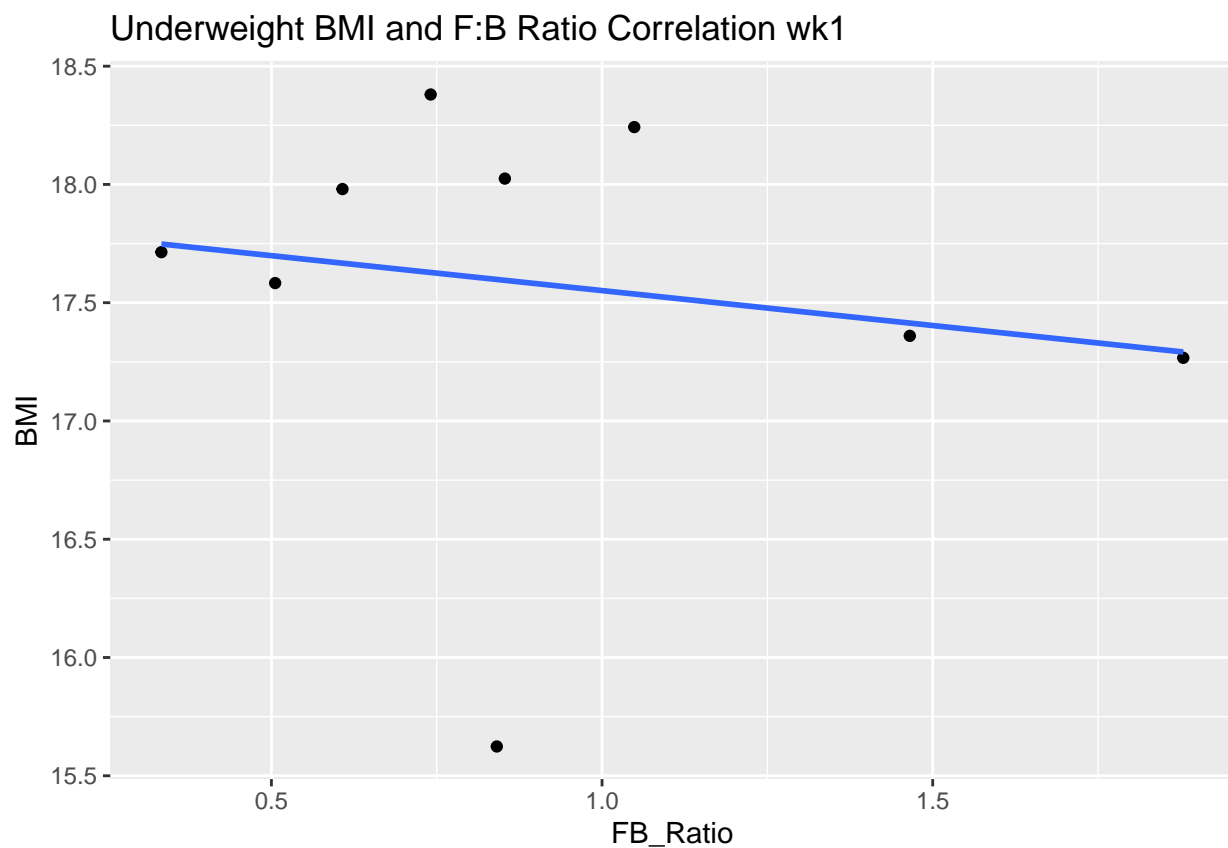
```
save_plot(filename = "figures/lm_under_BMI_plot_wk1.pdf",
          plot = lm_under_BMI_plot_wk1,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the BMI of "underweight" participants was 0.6539, which is not statistically significant. Furthermore, the R value was extremely low (-0.1082), and therefore insignificant. There does not appear to be a statistically significant linear relationship between the week 1 firmicutes to bacteroidetes ratio and BMI among participants whose BMI falls within the CDC's "underweight" category.

###F:B and Butyrate

```
lm_wk1_but_under_plot<-underweight_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = but_wk1)) +
  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("F:B ratio wk1") +
  ylab("Butyrate Concentration kg/mmol wk1")  + labs(title = "Underweight F:B Ratio and Butyrate Concen
print(lm_wk1_but_under_plot)
```


Underweight F:B Ratio and Butyrate Concentration Correlation wk1
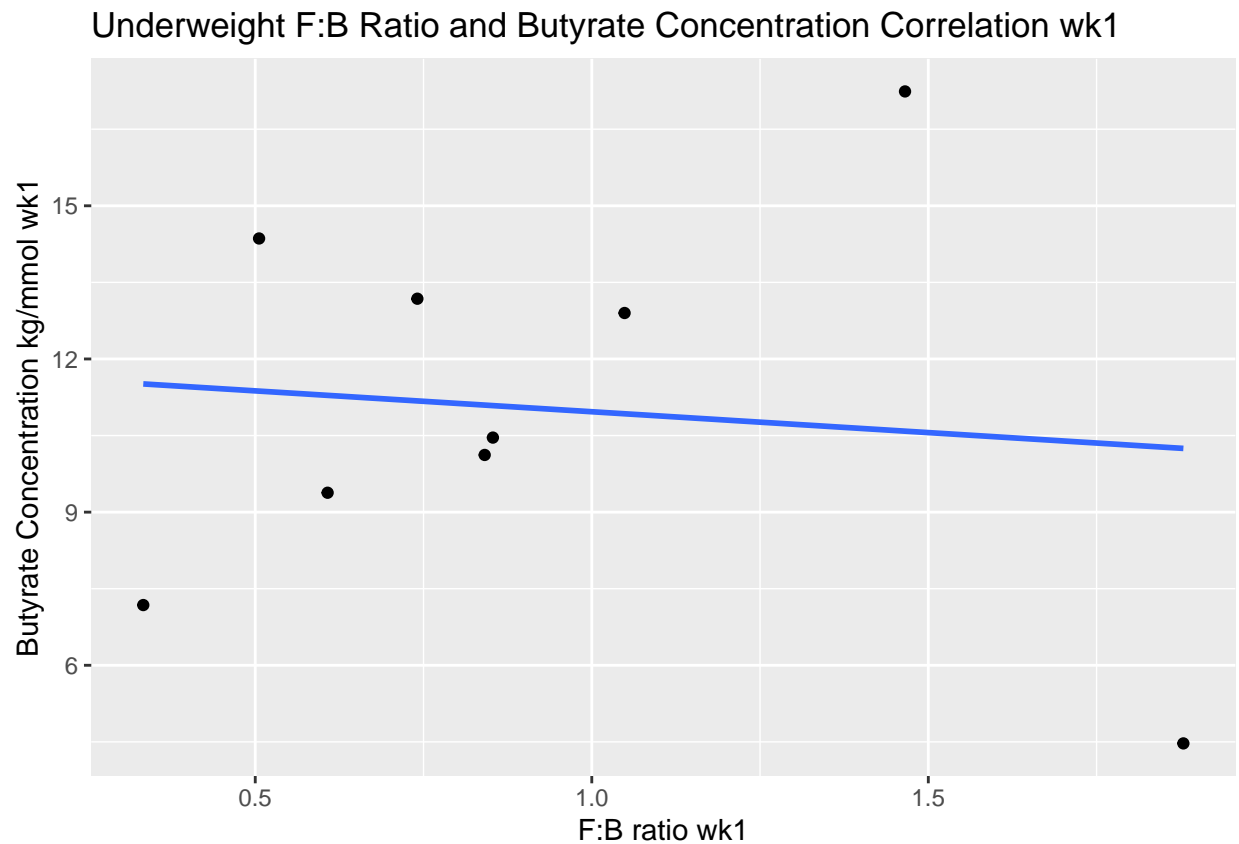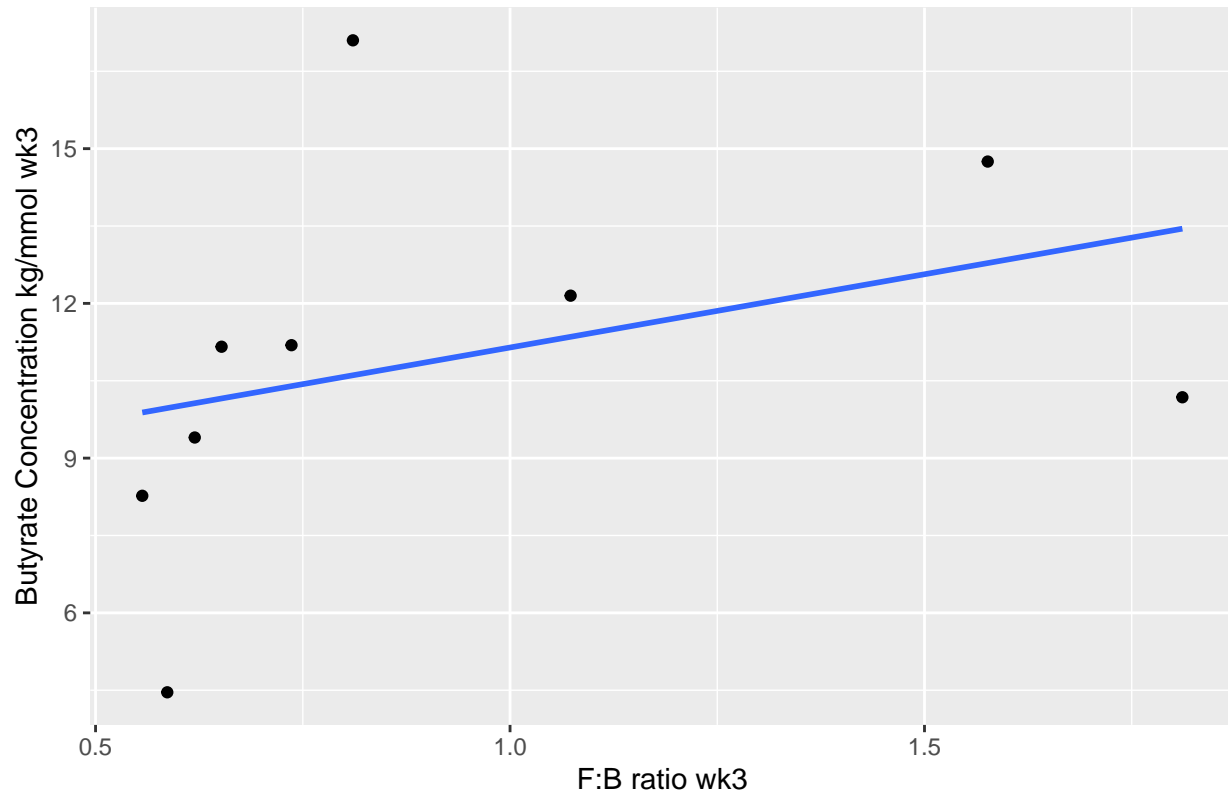
```r
lm_wk1_but_under<-underweight_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ but_wk1, data = .) #test relationship
summary(lm_wk1_but_under)
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ but_wk1, data = .)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.63556 -0.33327 -0.09032   0.15344   0.87472
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.06200    0.54926    1.933   0.0944 .
## but_wk1     -0.01292    0.04728   -0.273   0.7924
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5168 on 7 degrees of freedom
## Multiple R-squared:  0.01056,    Adjusted R-squared:  -0.1308
## F-statistic: 0.07474 on 1 and 7 DF,  p-value: 0.7924
```

```r
lm_wk3_but_under_plot<-underweight_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
             y = but_wk3)) +
  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("F:B ratio wk3") +
  ylab("Butyrate Concentration kg/mmol wk3")  + labs(title = "Underweight F:B Ratio and Butyrate Concen
print(lm_wk3_but_under_plot)
```

Underweight F:B Ratio and Butyrate Concentration Correlation wk3


```
lm_wk3_but_under<-underweight_phylum_scfa %>%
  lm(fb_ratio_wk3 ~ but_wk3, data = .) #test relationship
summary(lm_wk3_but_under)
```

```
##
## Call:
## lm(formula = fb_ratio_wk3 ~ but_wk3, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40413 -0.25710 -0.20974  0.08336  0.91089
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.43773    0.51248   0.854    0.421
## but_wk3      0.04543    0.04461   1.018    0.342
##
## Residual standard error: 0.4595 on 7 degrees of freedom
## Multiple R-squared:  0.129,  Adjusted R-squared:  0.004608
## F-statistic: 1.037 on 1 and 7 DF,  p-value: 0.3424
```

```
save_plot(filename = "figures/lm_wk1_but_under_plot.pdf",
          plot = lm_wk1_but_under_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

```
save_plot(filename = "figures/lm_wk3_but_under_plot.pdf",
          plot = lm_wk3_but_under_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the week 1 Butyrate concentration of the underweight participants is .79, which is not statistically significant. However, the R^2 value is insignificant being very small at R^2=-.1308, which represents too much variation in the data. There is no significant correlation between the week 1 firmicutes - bacteroidetes ratio and Butyrate Concentration.

The week 3 F:B ratio and week 3 Butyrate concentration of the underweight participants has a insignifcant p-value and R^2 value at p=.3424 and R^2 = 0.004608 respectively. This also represents no significant correlation.

###F:B and Propionate

```
lm_wk1_pro_under_plot<-underweight_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = pro_wk1)) +
  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("F:B ratio wk1") +
  ylab("Propionate Concentration kg/mmol wk1")  + labs(title = "Underweight F:B Ratio and Propionate Con
print(lm_wk1_pro_under_plot)
```


Underweight F:B Ratio and Propionate Concentration Correlation wk1
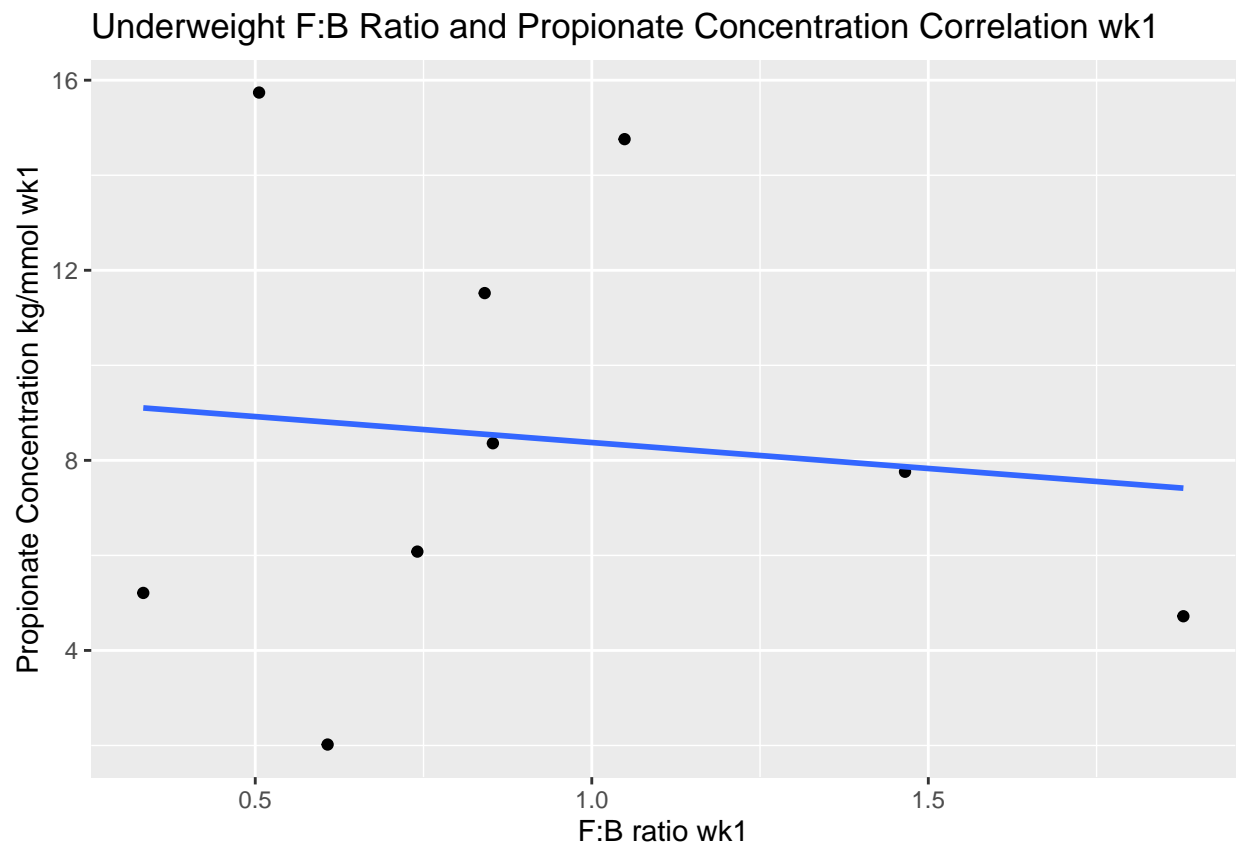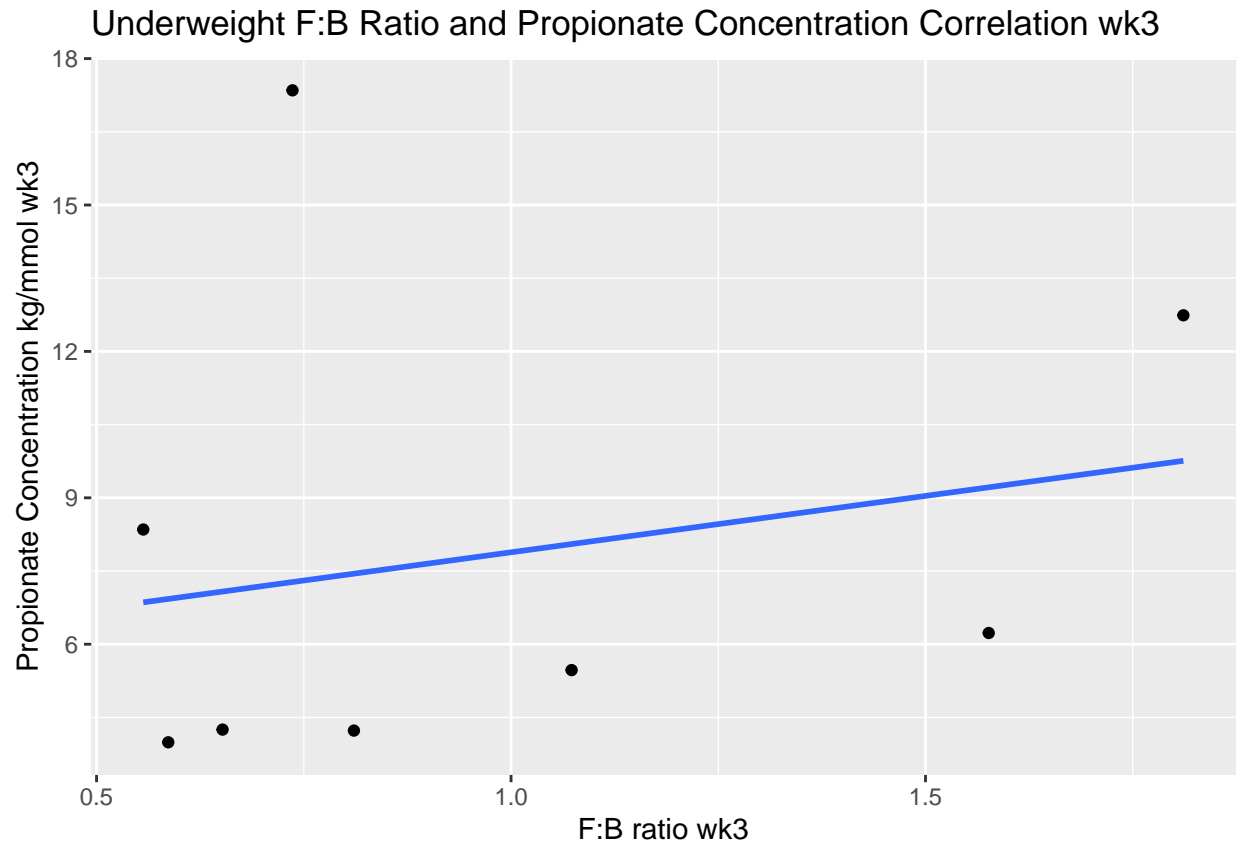
```
lm_wk1_pro_under<-underweight_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ pro_wk1, data = .) #test relationship
summary(lm_wk1_pro_under)
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ pro_wk1, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62432 -0.32751 -0.06757  0.20391  0.91518
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.01969    0.37301   2.734   0.0292 *
## pro_wk1     -0.01185    0.03910  -0.303   0.7707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5162 on 7 degrees of freedom
## Multiple R-squared:  0.01295,    Adjusted R-squared:  -0.1281
## F-statistic: 0.09181 on 1 and 7 DF,  p-value: 0.7707
```

```
lm_wk3_pro_under_plot<-underweight_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
             y = pro_wk3)) +
  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("F:B ratio wk3") +
  ylab("Propionate Concentration kg/mmol wk3")  + labs(title = "Underweight F:B Ratio and Propionate Co
print(lm_wk3_pro_under_plot)
```

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

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

Underweight F:B Ratio and Propionate Concentration Correlation wk3


```
lm_wk3_pro_under<-underweight_phylum_scfa %>%
  lm(fb_ratio_wk3 ~ pro_wk3, data = .) #test relationship
summary(lm_wk3_pro_under)
```

```
##
## Call:
## lm(formula = fb_ratio_wk3 ~ pro_wk3, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4521 -0.3348 -0.1638  0.2721  0.7259
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.80012    0.35362   2.263   0.0643 .
## pro_wk3      0.02238    0.03912   0.572   0.5880
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5005 on 6 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.05172,    Adjusted R-squared:  -0.1063
## F-statistic: 0.3273 on 1 and 6 DF,  p-value: 0.588
```

```
save_plot(filename = "figures/lm_wk1_pro_under_plot.pdf",
          plot = lm_wk1_pro_under_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)

save_plot(filename = "figures/lm_wk3_pro_under_plot.pdf",
          plot = lm_wk3_pro_under_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

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

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the week 1 Propionate concentration of all the participants is .7707, which is not statistically significant. Additionally, the R^2 value is insignificant being very small at R^2=-0.01281, which represents too much variation in the data. There is no significant correlation between the week 1 firmicutes - bacteroidetes ratio and Propionate concentration for underweight participants.

Similarly, The week 3 F - B ratio and week 3 Propionate concentration for underweight participants has a signifcant p-value and insignificant R^2 value at p=0.588 and R^2 = -0.1063 respectively. This also represents no significant correlation.

##Healthy Weight ###Wk1 F:B and BMI

```
#PLOT
hwk1fb_BMI_plot<-healthy_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = BMI)) +
  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
 labs(title = "Ratio of Firmicutes to Bacteroidetes by healthy BMI",
        x = "Ratio of Firmicutes to Bacteroidetes",
        y = "BMI")
print(hwk1fb_BMI_plot)
```

Ratio of Firmicutes to Bacteroidetes by healthy BMI
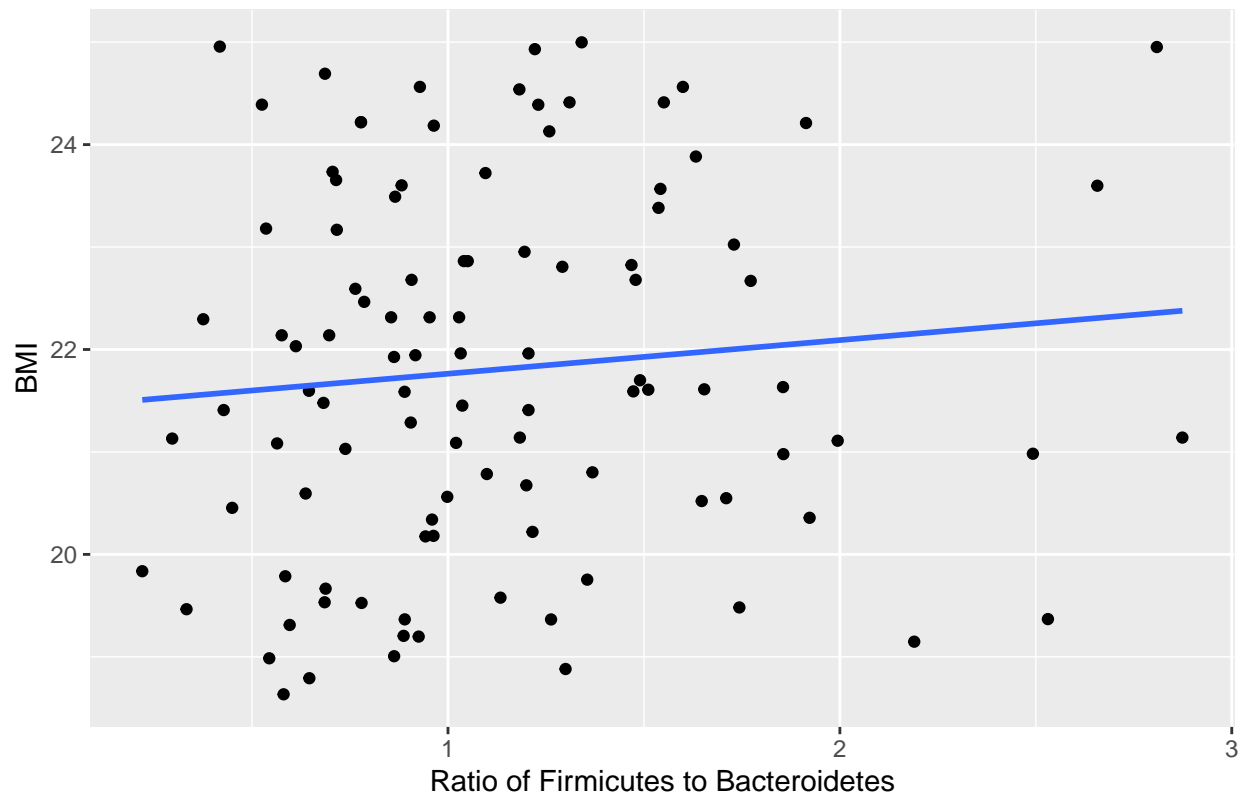

```
#TEST
hwk1fb_BMI_test <- healthy_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ BMI, data = .) #test relationship
summary(hwk1fb_BMI_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ BMI, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8493 -0.3908 -0.1168  0.3073  1.7647
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.45952    0.65672   0.700    0.486
## BMI          0.03073    0.03002   1.024    0.308
##
## Residual standard error: 0.5471 on 103 degrees of freedom
## Multiple R-squared:  0.01008,    Adjusted R-squared:  0.000465
## F-statistic: 1.048 on 1 and 103 DF,  p-value: 0.3083
```

```
#p-value = 0.3083
#Adjusted R-squared = 0.000465
```

```
#Save Plot
save_plot(filename = "figures/hwk1fb_BMI_plot.pdf",
          plot = hwk1fb_BMI_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the BMI of "healthy" participants was 0.3083, which is not statistically significant. Furthermore, the R value was an insignificant value close to zero (0.000465). There does not appear to be a statistically significant linear relationship between the week 1 firmicutes to bacteroidetes ratio and BMI among participants whose BMI falls within the CDC's "Healthy" category.

###Wk1 F:B and Wk1 Butyrate

```
#PLOT
hwk1fb_wk1but_plot<-healthy_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = but_wk1)) +
  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
 labs(title = "Week 1 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration",
       x = "Ratio of Firmicutes to Bacteroidetes",
       y = "Average Butyrate Concentration (mmol/kg)")
print(hwk1fb_wk1but_plot)
```


Week 1 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration
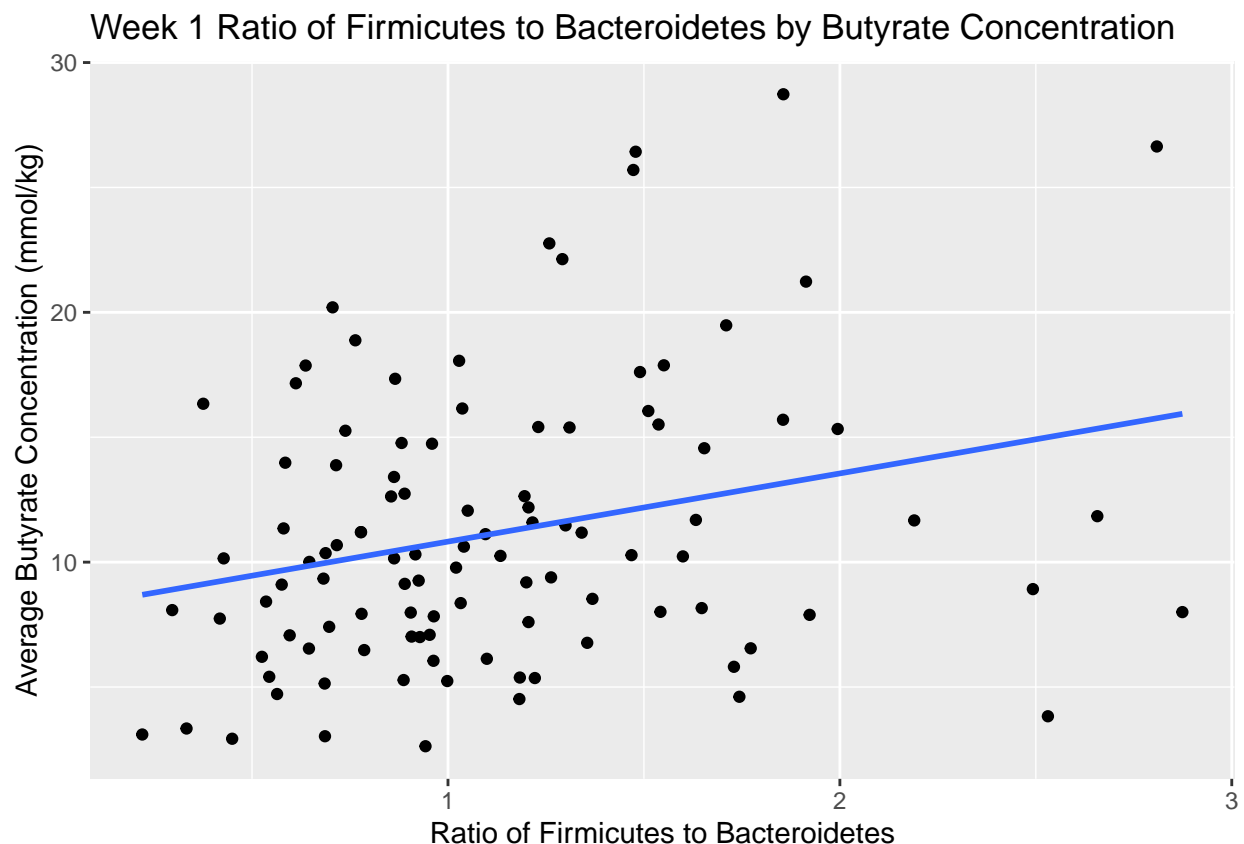
```
#TEST
hwk1fb_wk1but_test <- healthy_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ but_wk1, data = .) #test relationship
summary(hwk1fb_wk1but_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ but_wk1, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.89047 -0.36217 -0.07759  0.24425  1.82806
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.834786   0.116413   7.171 1.17e-10 ***
## but_wk1     0.026389   0.009333   2.828  0.00564 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5297 on 103 degrees of freedom
## Multiple R-squared:  0.07203,    Adjusted R-squared:  0.06302
## F-statistic: 7.995 on 1 and 103 DF,  p-value: 0.005637
```

```
#p-value = 0.005637
#Adjusted R-squared = 0.06302

#Save Plot
save_plot(filename = "figures/hwk1fb_wk1but_plot.pdf",
          plot = hwk1fb_wk1but_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the week 1 average butyrate concentration among "healthy" participants was 0.005637, which falls within a 95% confidence interval. However, the R value was an insignificant value close to zero (0.06302). There does not appear to be a statistically significant linear relationship between the week 1 firmicutes to bacteroidetes ratio and the week 1 average butyrate concentration among participants whose BMI falls within the CDC's "Healthy" category.

###Wk1 F:B and Wk1 Propionate
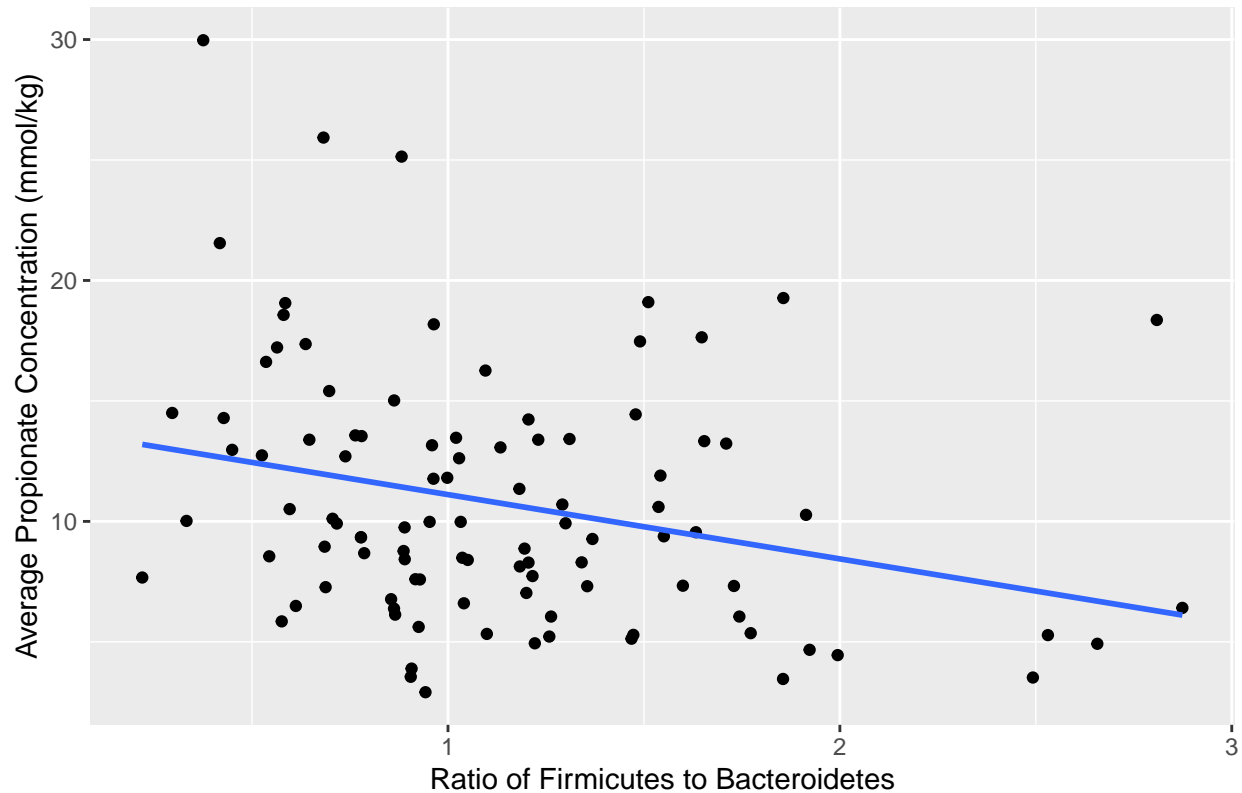
```
#PLOT
hwk1fb_wk1pro_plot<-healthy_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = pro_wk1)) +
  geom_point() + #puts data points to match x and y coordinates
  geom_smooth(method = "lm", #used to create a linear best fit line
              se = FALSE) + #hides confidence interval around line
 labs(title = "Week 1 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration",
        x = "Ratio of Firmicutes to Bacteroidetes",
        y = "Average Propionate Concentration (mmol/kg)")
print(hwk1fb_wk1pro_plot)
```

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

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

Week 1 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration


```
#TEST
hwk1fb_wk1pro_test <- healthy_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ pro_wk1, data = .) #test relationship
summary(hwk1fb_wk1pro_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ pro_wk1, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0026 -0.3938 -0.1033  0.1928  1.8976
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.44580    0.12032  12.017  < 2e-16 ***
## pro_wk1     -0.02912    0.01008  -2.889  0.00475 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5235 on 99 degrees of freedom
```

```
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.07773,    Adjusted R-squared:  0.06841
## F-statistic: 8.344 on 1 and 99 DF,  p-value: 0.004754
```

```
#p-value = 0.004754
#Adjusted R-squared = 0.06841

#Save Plot
save_plot(filename = "figures/hwk1fb_wk1pro_plot.pdf",
          plot = hwk1fb_wk1pro_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

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

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the week 1 average propionate concentration among "healthy" participants was 0.004754, which falls within a 95% confidence interval. However, the R value was an insignificant value close to zero (0.06841). There does not appear to be a statistically significant linear relationship between the week 1 firmicutes to bacteroidetes ratio and the week 1 average propionate concentration among participants whose BMI falls within the CDC's "Healthy" category.

Wk3 F:B and Wk3 Butyrate

```
#PLOT
hwk3fb_wk3but_plot<-healthy_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
             y = but_wk3)) +
  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
 labs(title = "Week 3 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration",
        x = "Ratio of Firmicutes to Bacteroidetes",
        y = "Average Butyrate Concentration (mmol/kg)")
print(hwk3fb_wk3but_plot)
```

Week 3 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration
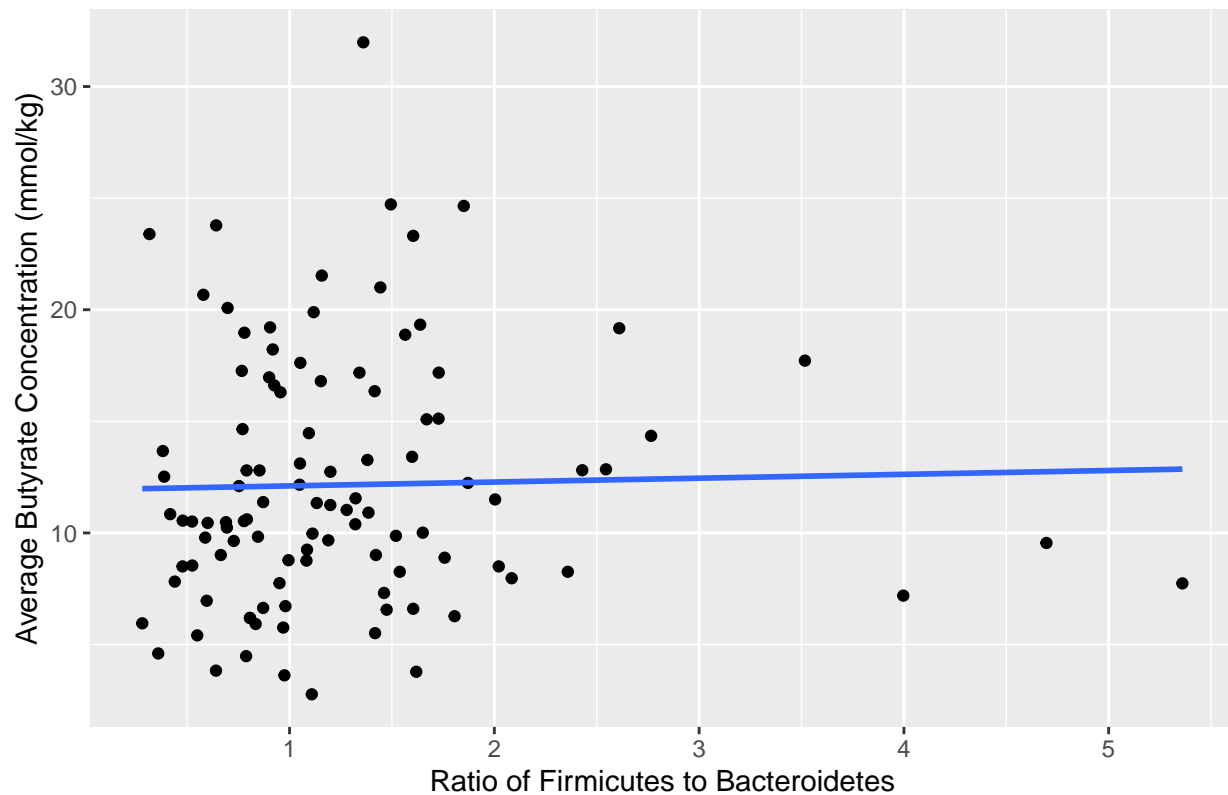

```
#TEST
hwk3fb_wk3but_test <- healthy_phylum_scfa %>%
  lm(fb_ratio_wk3 ~ but_wk3, data = .) #test relationship
summary(hwk3fb_wk3but_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk3 ~ but_wk3, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9964 -0.5083 -0.1810  0.2699  4.1076
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.222715   0.195979   6.239 9.89e-09 ***
## but_wk3     0.003825   0.014673   0.261    0.795
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8328 on 103 degrees of freedom
## Multiple R-squared:  0.0006594,  Adjusted R-squared:  -0.009043
## F-statistic: 0.06797 on 1 and 103 DF,  p-value: 0.7948
```

```
#p-value = 0.7948
#Adjusted R-squared = -0.009043

#Save Plot
save_plot(filename = "figures/hwk3fb_wk3but_plot.pdf",
          plot = hwk3fb_wk3but_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

The p-value for the linear correlation of the week 3 firmicutes to bacteroidetes ratio and the week 3 average butyrate concentration among "healthy" participants was 0.7948, which is not statistically significant. Furthermore, the R value was an insignificant value close to zero (-0.009043). There does not appear to be a statistically significant linear relationship between the week 3 firmicutes to bacteroidetes ratio and the week 3 average butyrate concentration among participants whose BMI falls within the CDC's "Healthy" category.

###Wk3 F:B and Wk3 Propionate

```
#PLOT
hwk3fb_wk3pro_plot<-healthy_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
             y = pro_wk3)) +
  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
 labs(title = "Week 3 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration",
         x = "Ratio of Firmicutes to Bacteroidetes",
         y = "Average Propionate Concentration (mmol/kg)")
print(hwk3fb_wk3pro_plot)
```

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

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

Week 3 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration
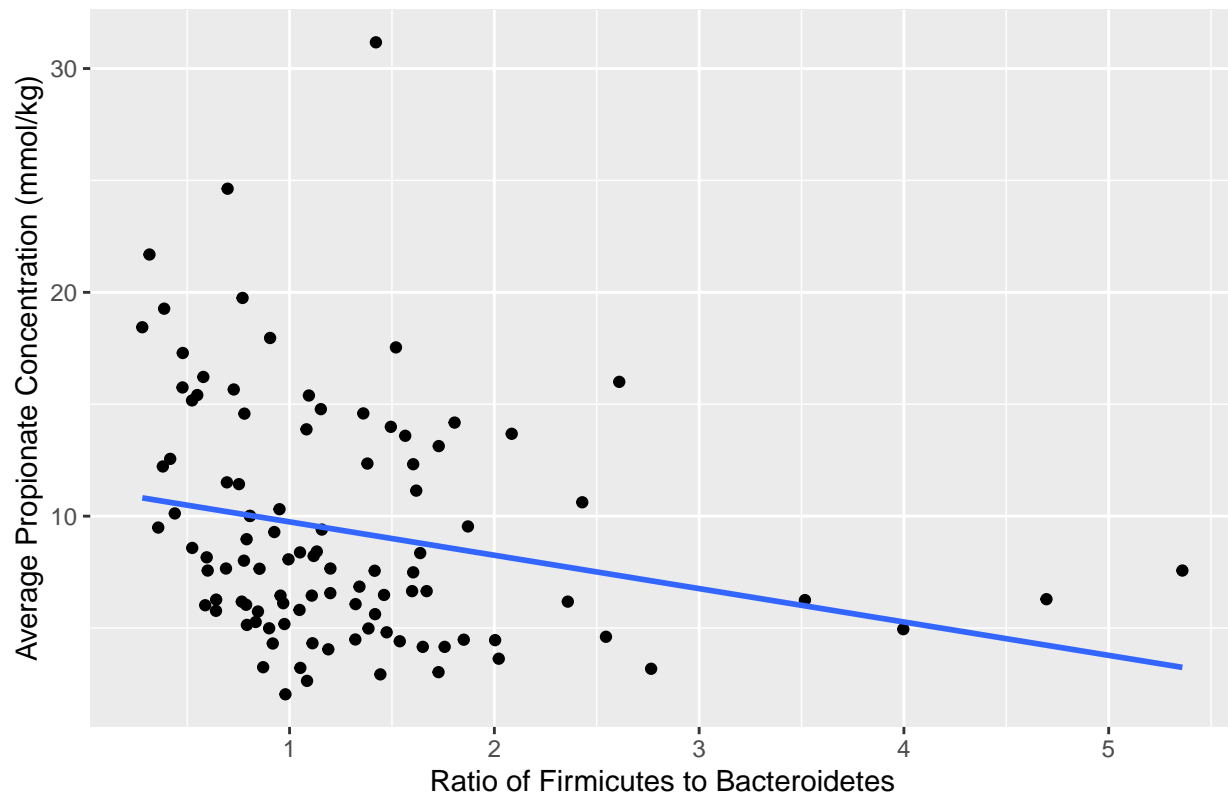
```
#TEST
hwk3fb_wk3pro_test <- healthy_phylum_scfa %>%
  lm(fb_ratio_wk3 ~ pro_wk3, data = .) #test relationship
summary(hwk3fb_wk3pro_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk3 ~ pro_wk3, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9143 -0.5030 -0.1932  0.2727  4.0151
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.62858    0.16498   9.871   <2e-16 ***
## pro_wk3     -0.03748    0.01540  -2.433   0.0167 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8185 on 100 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.05589,    Adjusted R-squared:  0.04645
## F-statistic:  5.92 on 1 and 100 DF,  p-value: 0.01675
```

```
#p-value = 0.01675
#Adjusted R-squared = 0.04645

#Save Plot
save_plot(filename = "figures/hwk3fb_wk3pro_plot.pdf",
          plot = hwk3fb_wk3pro_plot,
          nrow = 1, ncol = 1,
          base_aspect_ratio = 2.1)
```

Warning: Removed 3 rows containing non-finite values (stat_smooth).

Warning: Removed 3 rows containing missing values (geom_point).

The p-value for the linear correlation of the week 3 firmicutes to bacteroidetes ratio and the week 3 average propionate concentration among "healthy" participants was 0.01675, which falls within a 95% confidence interval. However, the R value was an insignificant value close to zero (0.04645). There does not appear to be a statistically significant linear relationship between the week 3 firmicutes to bacteroidetes ratio and the week 3 average propionate concentration among participants whose BMI falls within the CDC's "Healthy" category.

##Overweight ##Wk1 F:B and BMI

```
#PLOT
ovwk1fb_BMI_plot<-overweight_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = BMI)) +
  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
 labs(title = "Ratio of Firmicutes to Bacteroidetes by Overweight BMI",
        x = "Ratio of Firmicutes to Bacteroidetes",
        y = "BMI")
print(ovwk1fb_BMI_plot)
```

Ratio of Firmicutes to Bacteroidetes by Overweight BMI
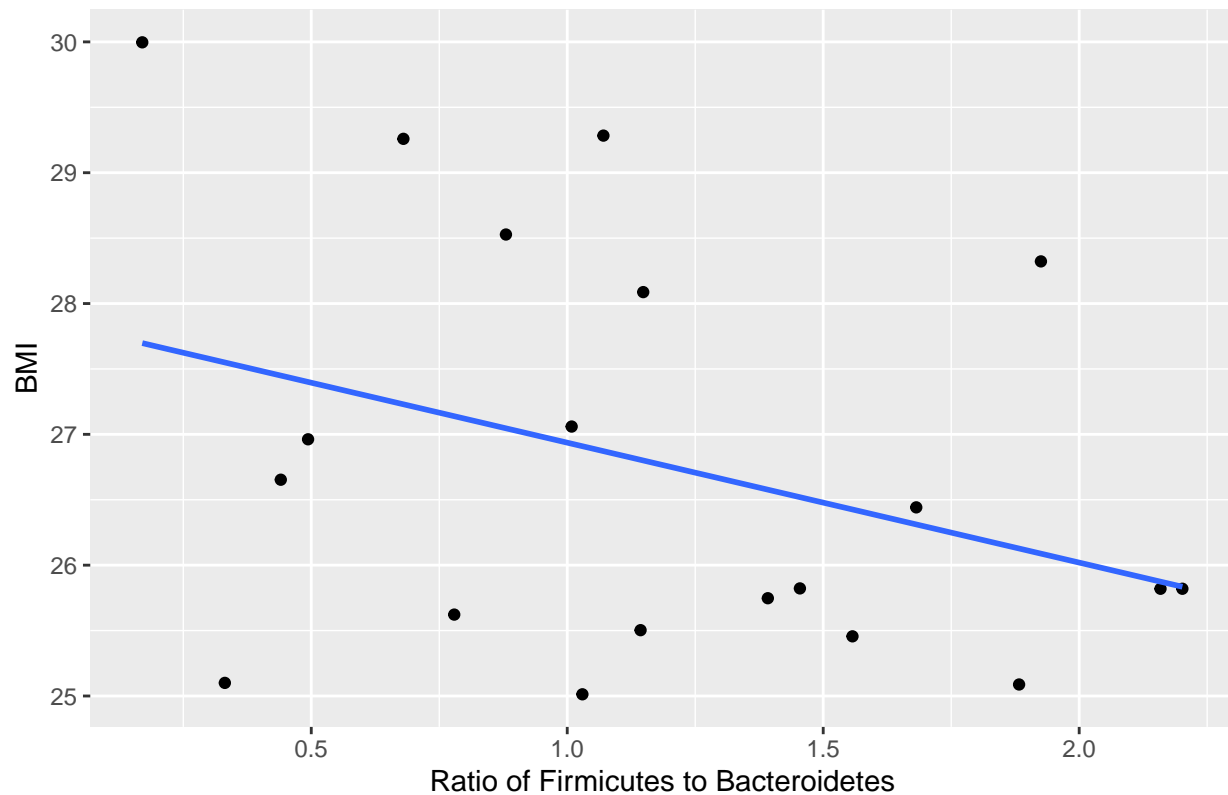

```
#TEST
ovwk1fb_BMI_test <- overweight_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ BMI, data = .) #test relationship
summary(ovwk1fb_BMI_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ BMI, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06352 -0.41901  0.01222  0.29048  0.95898
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.73184    2.25636   2.097   0.0504 .
## BMI         -0.13295    0.08412  -1.581   0.1314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5797 on 18 degrees of freedom
## Multiple R-squared:  0.1219, Adjusted R-squared:  0.07309
## F-statistic: 2.498 on 1 and 18 DF,  p-value: 0.1314
```

```
#p-value = 0.1314
#Adjusted R-squared = 0.07309

#Save Plot
save_plot(filename = "figures/ovwk1fb_BMI_plot.pdf", plot = ovwk1fb_BMI_plot, nrow = 1, ncol = 1, base_a
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the BMI of "overweight" participants was 0.1314, which is not statistically significant. Furthermore, the R value was an insignificant value close to zero (0.07309). There does not appear to be a statistically significant linear relationship between the week 1 firmicutes to bacteroidetes ratio and BMI among participants whose BMI falls within the CDC's "Overweight" category.

###Wk1 F:B and Wk1 Butyrate

```
#PLOT
ovwk1fb_wk1but_plot<-overweight_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = but_wk1)) +
  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
 labs(title = "Week 1 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration",
       x = "Ratio of Firmicutes to Bacteroidetes",
       y = "Average Butyrate Concentration (mmol/kg)")
print(ovwk1fb_wk1but_plot)
```


Week 1 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration
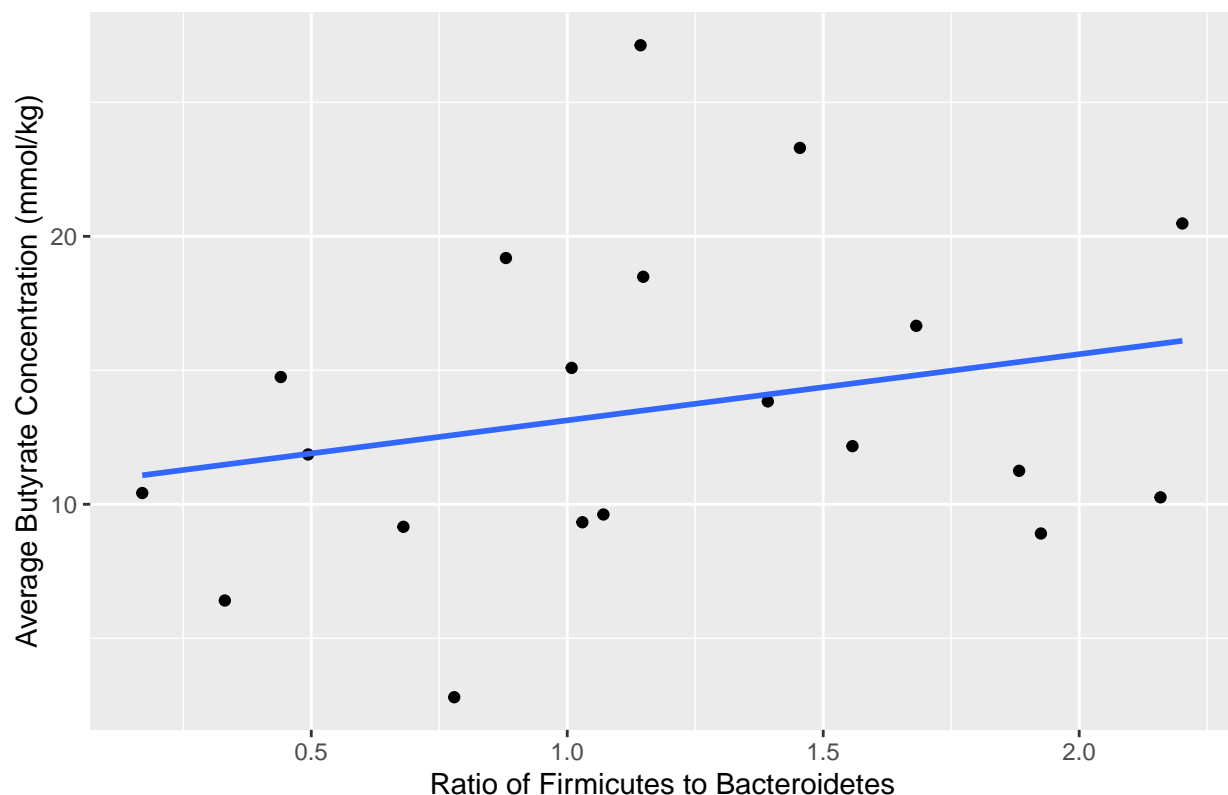
```
#TEST
ovwk1fb_wk1but_test <- overweight_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ but_wk1, data = .) #test relationship
summary(ovwk1fb_wk1but_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ but_wk1, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.92175 -0.39333 -0.07635  0.42356  1.07132
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.82642    0.34136   2.421   0.0263 *
## but_wk1      0.02545    0.02316   1.099   0.2863
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5989 on 18 degrees of freedom
## Multiple R-squared:  0.06286,    Adjusted R-squared:  0.0108
## F-statistic: 1.207 on 1 and 18 DF,  p-value: 0.2863
```

```
#p-value = 0.2863
#Adjusted R-squared = 0.0108

#Save Plot
save_plot(filename = "figures/ovwk1fb_wk1but_plot.pdf", plot = ovwk1fb_wk1but_plot, nrow = 1, ncol = 1,
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the week 1 average butyrate concentration among "overweight" participants was 0.2863, which falls within a 95% confidence interval. However, the R value was an insignificant value close to zero (0.0108). There does not appear to be a statistically significant linear relationship between the week 1 firmicutes to bacteroidetes ratio and the week 1 average butyrate concentration among participants whose BMI falls within the CDC's "Overweight" category.

Wk1 F:B and Wk1 Propionate
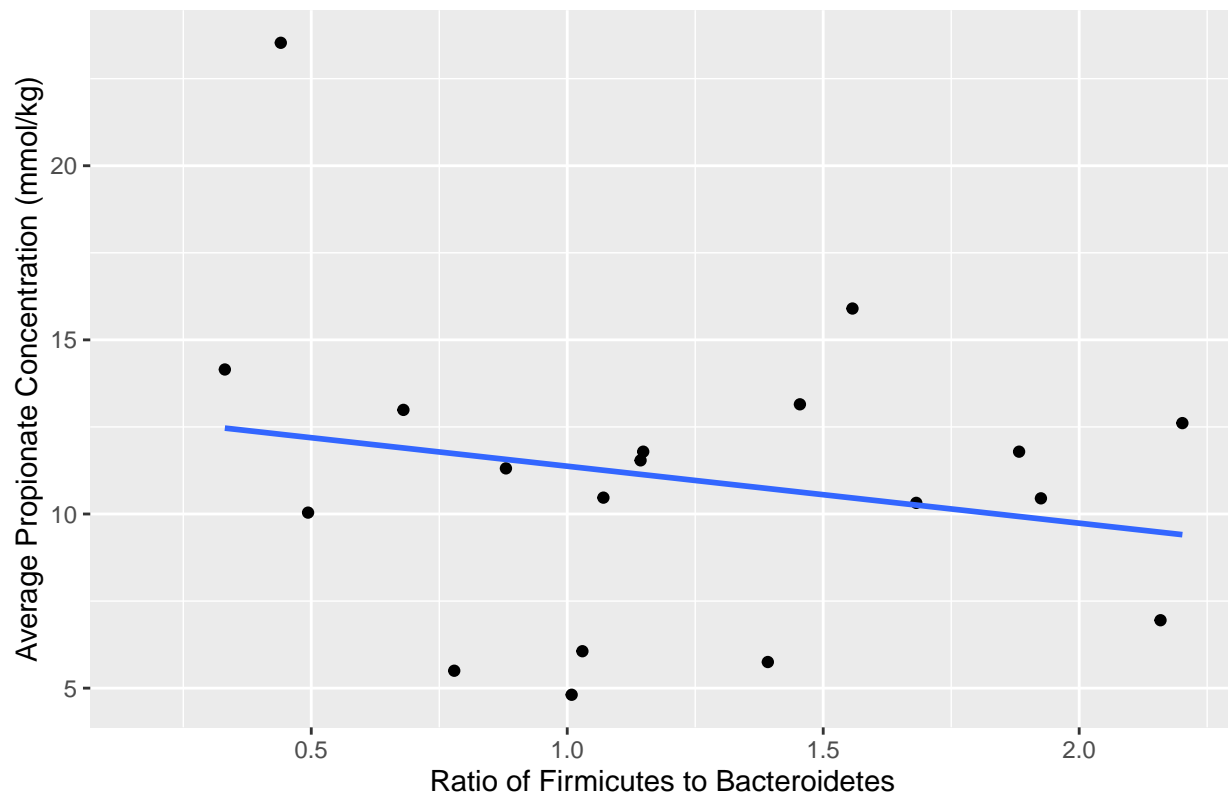
```
#PLOT
ovwk1fb_wk1pro_plot<-overweight_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = pro_wk1)) +
  geom_point() + #puts data points to match x and y coordinates
  geom_smooth(method = "lm", #used to create a linear best fit line
              se = FALSE) + #hides confidence interval around line
 labs(title = "Week 1 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration",
       x = "Ratio of Firmicutes to Bacteroidetes",
       y = "Average Propionate Concentration (mmol/kg)")
print(ovwk1fb_wk1pro_plot)
```

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

Warning: Removed 1 rows containing missing values (geom_point).

Week 1 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration


```
#TEST
ovwk1fb_wk1pro_test <- overweight_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ pro_wk1, data = .) #test relationship
summary(ovwk1fb_wk1pro_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ pro_wk1, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.80543 -0.41150 -0.06605  0.45390  1.02211
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.53069    0.36471   4.197 0.000605 ***
## pro_wk1     -0.02785    0.03092  -0.901 0.380231
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5723 on 17 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.04557,    Adjusted R-squared:  -0.01058
## F-statistic: 0.8116 on 1 and 17 DF,  p-value: 0.3802
```

```
#p-value = 0.3802
#Adjusted R-squared = -0.01058

#Save Plot
save_plot(filename = "figures/ovwk1fb_wk1pro_plot.pdf", plot = ovwk1fb_wk1pro_plot, nrow = 1, ncol = 1,
```

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

Warning: Removed 1 rows containing missing values (geom_point).

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the week 1 average propionate concentration among "overweight" participants was 0.3802, which falls within a 95% confidence interval. However, the R value was an insignificant value close to zero (-0.01058). There does not appear to be a statistically significant linear relationship between the week 1 firmicutes to bacteroidetes ratio and the week 1 average propionate concentration among participants whose BMI falls within the CDC's "Overweight" category.

Wk3 F:B and Wk3 Butyrate

```
#PLOT
ovwk3fb_wk3but_plot<-overweight_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
             y = but_wk3)) +
  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
 labs(title = "Week 3 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration",
        x = "Ratio of Firmicutes to Bacteroidetes",
        y = "Average Butyrate Concentration (mmol/kg)")
print(ovwk3fb_wk3but_plot)
```

Week 3 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration
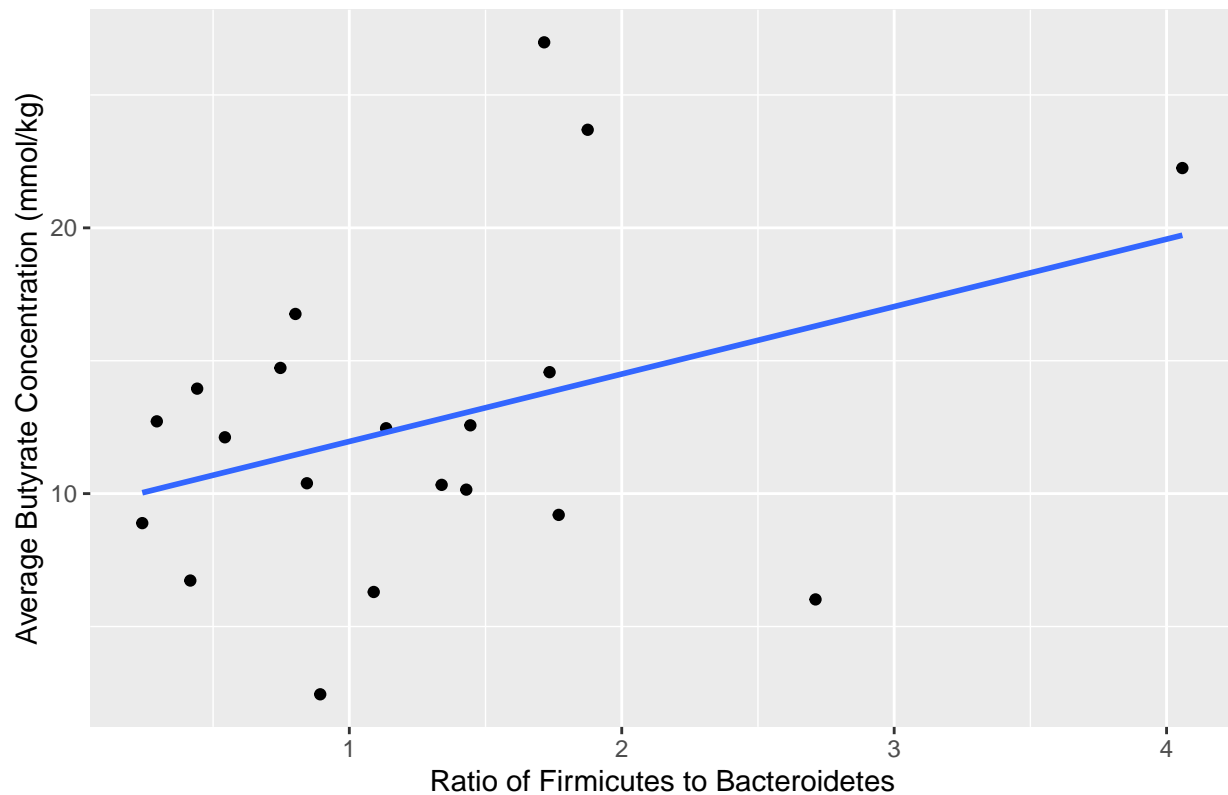

```
#TEST
ovwk3fb_wk3but_test <- overweight_phylum_scfa %>%
  lm(fb_ratio_wk3 ~ but_wk3, data = .) #test relationship
summary(ovwk3fb_wk3but_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk3 ~ but_wk3, data = .)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -0.98589 -0.65940 -0.07489  0.21897  2.24361
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.56514    0.45450   1.243    0.230
## but_wk3      0.05615    0.03247   1.729    0.101
##
## Residual standard error: 0.8666 on 18 degrees of freedom
## Multiple R-squared:  0.1425, Adjusted R-squared:  0.09483
## F-statistic: 2.991 on 1 and 18 DF,  p-value: 0.1009
```

```
#p-value = 0.1009
#Adjusted R-squared = 0.09483
```

```
#Save Plot
save_plot(filename = "figures/ovwk3fb_wk3but_plot.pdf", plot = ovwk3fb_wk3but_plot, nrow = 1, ncol = 1,
```

The p-value for the linear correlation of the week 3 firmicutes to bacteroidetes ratio and the week 3 average butyrate concentration among "overweight" participants was 0.1009, which is not statistically significant. Furthermore, the R value was an insignificant value close to zero (0.09483). There does not appear to be a statistically significant linear relationship between the week 3 firmicutes to bacteroidetes ratio and the week 3 average butyrate concentration among participants whose BMI falls within the CDC's "Overweight" category.

Wk3 F:B and Wk3 Propionate

```
#PLOT
ovwk3fb_wk3pro_plot<-overweight_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
             y = pro_wk3)) +
  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
 labs(title = "Week 3 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration",
         x = "Ratio of Firmicutes to Bacteroidetes",
         y = "Average Propionate Concentration (mmol/kg)")
print(ovwk3fb_wk3pro_plot)
```


Week 3 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration
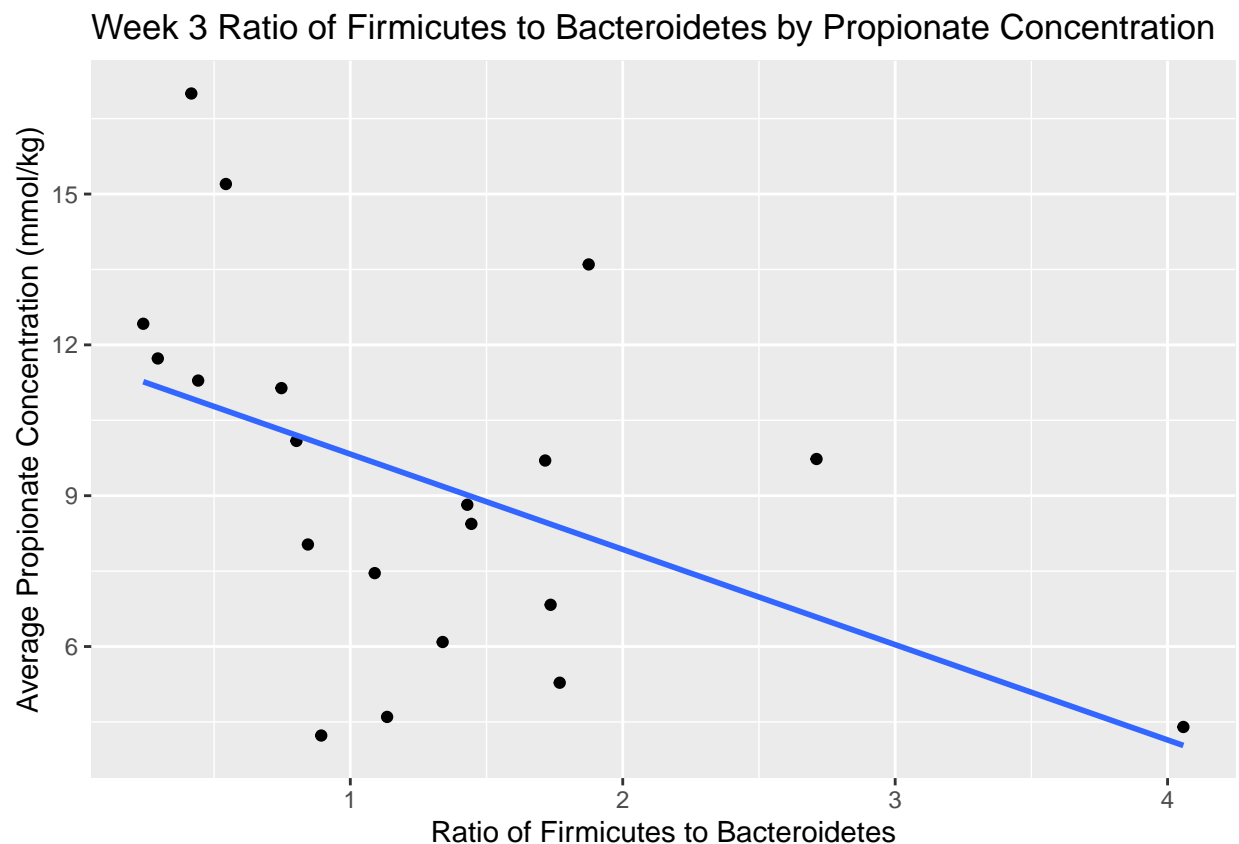
```
#TEST
ovwk3fb_wk3pro_test <- overweight_phylum_scfa %>%
  lm(fb_ratio_wk3 ~ pro_wk3, data = .) #test relationship
summary(ovwk3fb_wk3pro_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk3 ~ pro_wk3, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9969 -0.5880 -0.1631  0.1110  2.1885
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.40195    0.52000   4.619 0.000213 ***
## pro_wk3     -0.12101    0.05229  -2.314 0.032668 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8215 on 18 degrees of freedom
## Multiple R-squared:  0.2293, Adjusted R-squared:  0.1865
## F-statistic: 5.356 on 1 and 18 DF,  p-value: 0.03267
```

```
#p-value = 0.03267
#Adjusted R-squared = 0.1865

#Save Plot
save_plot(filename = "figures/ovwk3fb_wk3pro_plot.pdf", plot = ovwk3fb_wk3pro_plot, nrow = 1, ncol = 1,
```

The p-value for the linear correlation of the week 3 firmicutes to bacteroidetes ratio and the week 3 average propionate concentration among "healthy" participants was 0.03267, which technically supports the alternative hypothesis. However, the R value was an insignificant value (0.1865). There does not appear to be a statistically significant linear relationship between the week 3 firmicutes to bacteroidetes ratio and the week 3 average propionate concentration among participants whose BMI falls within the CDC's "Overweight" category.

##Obese ###Wk1 F:B and BMI

```
#PLOT
obwk1fb_BMI_plot<-obese_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = BMI)) +
  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
 labs(title = "Ratio of Firmicutes to Bacteroidetes by Obese BMI",
        x = "Ratio of Firmicutes to Bacteroidetes",
        y = "BMI")
print(obwk1fb_BMI_plot)
```

Ratio of Firmicutes to Bacteroidetes by Obese BMI
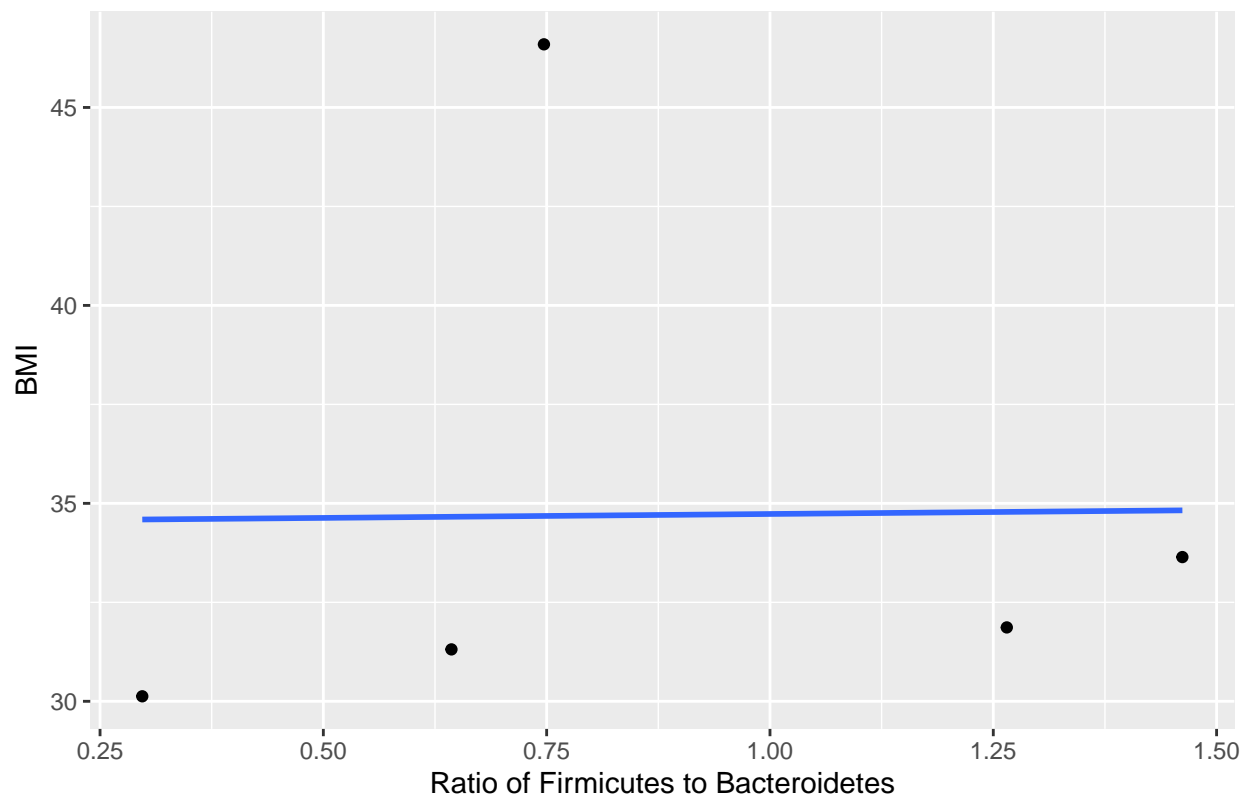

```
#TEST
obwk1fb_BMI_test <- obese_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ BMI, data = .) #test relationship
summary(obwk1fb_BMI_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ BMI, data = .)
##
## Residuals:
##       1       2       3       4       5
## -0.2362 -0.1475  0.5798 -0.5811  0.3850
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.8491209  1.4261369   0.595    0.594
## BMI         0.0009733  0.0404790   0.024    0.982
##
## Residual standard error: 0.5476 on 3 degrees of freedom
## Multiple R-squared:  0.0001927,  Adjusted R-squared:  -0.3331
## F-statistic: 0.0005781 on 1 and 3 DF,  p-value: 0.9823
```

```
#p-value = 0.9823
#Adjusted R-squared = 0.3331
```

```
#Save Plot
save_plot(filename = "figures/obwk1fb_BMI_plot.pdf", plot = obwk1fb_BMI_plot, nrow = 1, ncol = 1, base_a
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the BMI of "obese" participants was 0.9823, which is not statistically significant. Furthermore, the R value was an insignificant value (0.3331). There does not appear to be a statistically significant linear relationship between the week 1 firmicutes to bacteroidetes ratio and BMI among participants whose BMI falls within the CDC's "Obese" category.

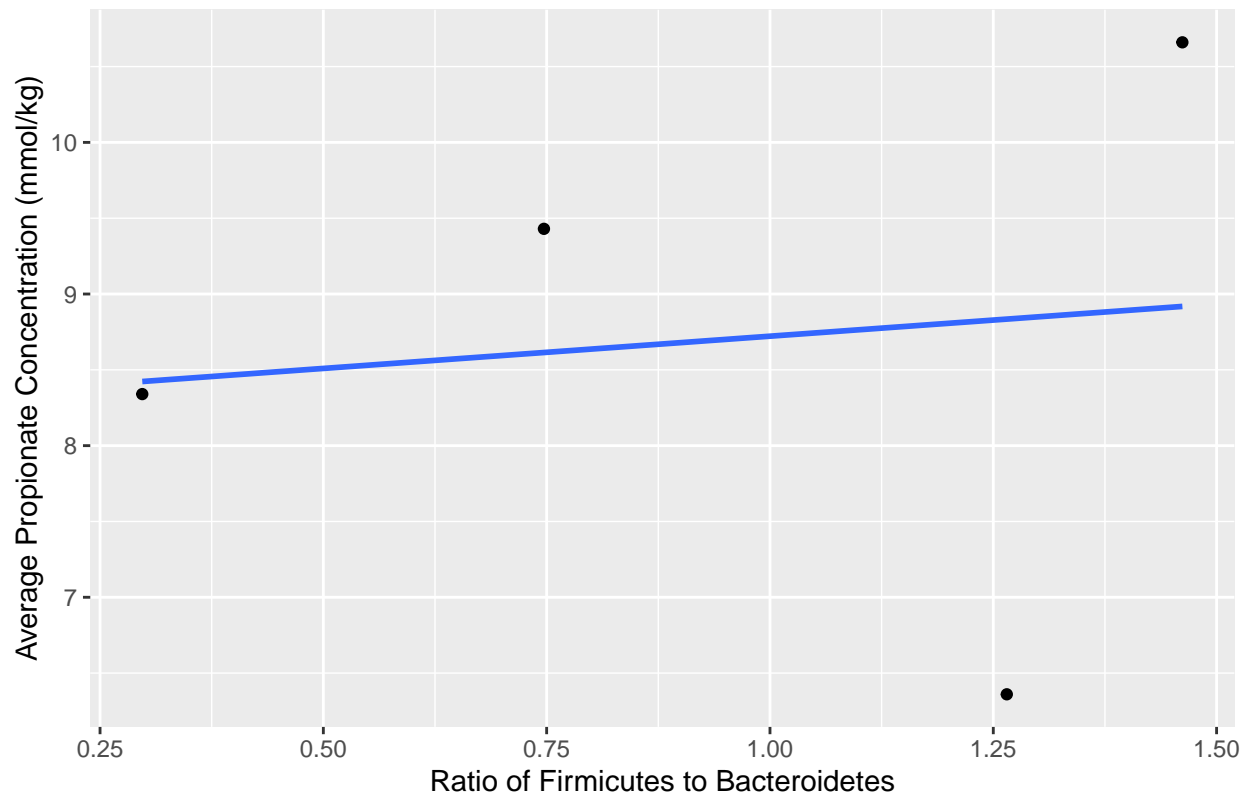### Wk1 F:B and Wk1 Butyrate

```
#PLOT
obwk1fb_wk1but_plot<-obese_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = but_wk1)) +
  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
 labs(title = "Week 1 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration",
         x = "Ratio of Firmicutes to Bacteroidetes",
         y = "Average Butyrate Concentration (mmol/kg)")
print(obwk1fb_wk1but_plot)
```


Week 1 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration
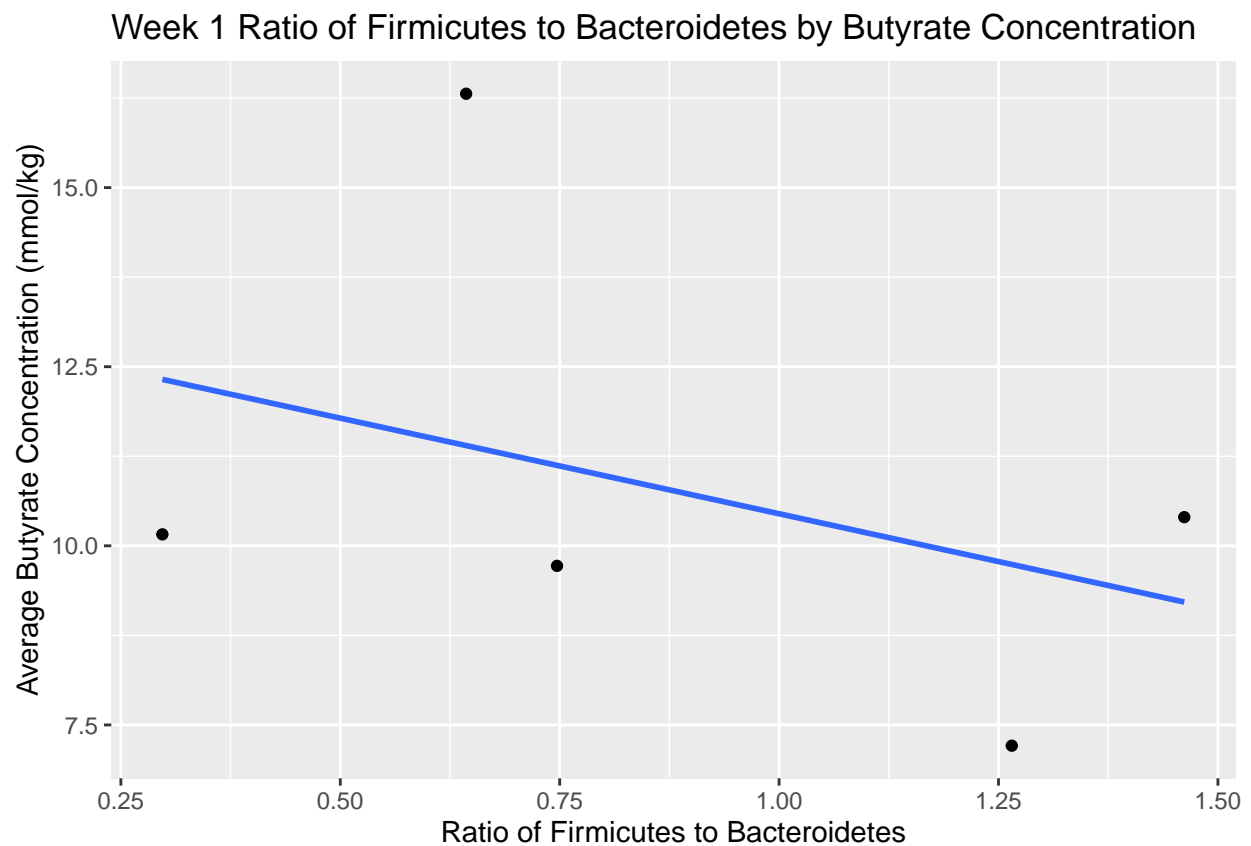
```
#TEST
obwk1fb_wk1but_test <- obese_phylum_scfa %>%
```

```
   lm(fb_ratio_wk1 ~ but_wk1, data = .) #test relationship
summary(obwk1fb_wk1but_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ but_wk1, data = .)
##
## Residuals:
##         1        2        3        4        5
##   0.05688 -0.19146  0.55956 -0.61765  0.19267
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.45749    0.84471   1.725    0.183
## but_wk1     -0.05340    0.07562  -0.706    0.531
##
## Residual standard error: 0.5072 on 3 degrees of freedom
## Multiple R-squared:  0.1425, Adjusted R-squared:  -0.1433
## F-statistic: 0.4986 on 1 and 3 DF,  p-value: 0.531
```

```
#p-value = 0.531
#Adjusted R-squared = -0.1433

#Save Plot
save_plot(filename = "figures/obwk1fb_wk1but_plot.pdf", plot = obwk1fb_wk1but_plot, nrow = 1, ncol = 1,
```

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the week 1 average butyrate concentration among "obese" participants was 0.531, which falls within a 95% confidence interval. However, the R value was an insignificant value (-0.1433). There does not appear to be a statistically significant linear relationship between the week 1 firmicutes to bacteroidetes ratio and the week 1 average butyrate concentration among participants whose BMI falls within the CDC's "Obese" category.

Wk1 F:B and Wk1 Propionate

```
#PLOT
obwk1fb_wk1pro_plot<-obese_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk1,
             y = pro_wk1)) +
  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
 labs(title = "Week 1 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration",
       x = "Ratio of Firmicutes to Bacteroidetes",
       y = "Average Propionate Concentration (mmol/kg)")
print(obwk1fb_wk1pro_plot)
```

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

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

Week 1 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration


```
#TEST
obwk1fb_wk1pro_test <- obese_phylum_scfa %>%
  lm(fb_ratio_wk1 ~ pro_wk1, data = .) #test relationship
summary(obwk1fb_wk1pro_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk1 ~ pro_wk1, data = .)
##
## Residuals:
##        2       3       4       5
## -0.2216  0.4497 -0.6329  0.4049
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.63585    1.78708   0.356    0.756
## pro_wk1      0.03529    0.20216   0.175    0.877
##
## Residual standard error: 0.6386 on 2 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.01501,    Adjusted R-squared:  -0.4775
## F-statistic: 0.03047 on 1 and 2 DF,  p-value: 0.8775
```

```
#p-value = 0.8775
#Adjusted R-squared = -0.4775
```

```
#Save Plot
save_plot(filename = "figures/obwk1fb_wk1pro_plot.pdf", plot = obwk1fb_wk1pro_plot, nrow = 1, ncol = 1,
```

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

Warning: Removed 1 rows containing missing values (geom_point).

The p-value for the linear correlation of the week 1 firmicutes to bacteroidetes ratio and the week 1 average propionate concentration among "obese" participants was 0.8775, which falls within a 95% confidence interval. However, the R value was an insignificant value (-0.4775). There does not appear to be a statistically significant linear relationship between the week 1 firmicutes to bacteroidetes ratio and the week 1 average propionate concentration among participants whose BMI falls within the CDC's "Obese" category.

###Wk3 F:B and Wk3 Butyrate

```
#PLOT
obwk3fb_wk3but_plot<-obese_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
             y = but_wk3)) +
  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
 labs(title = "Week 3 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration",
        x = "Ratio of Firmicutes to Bacteroidetes",
        y = "Average Butyrate Concentration (mmol/kg)")
print(obwk3fb_wk3but_plot)
```

Week 3 Ratio of Firmicutes to Bacteroidetes by Butyrate Concentration
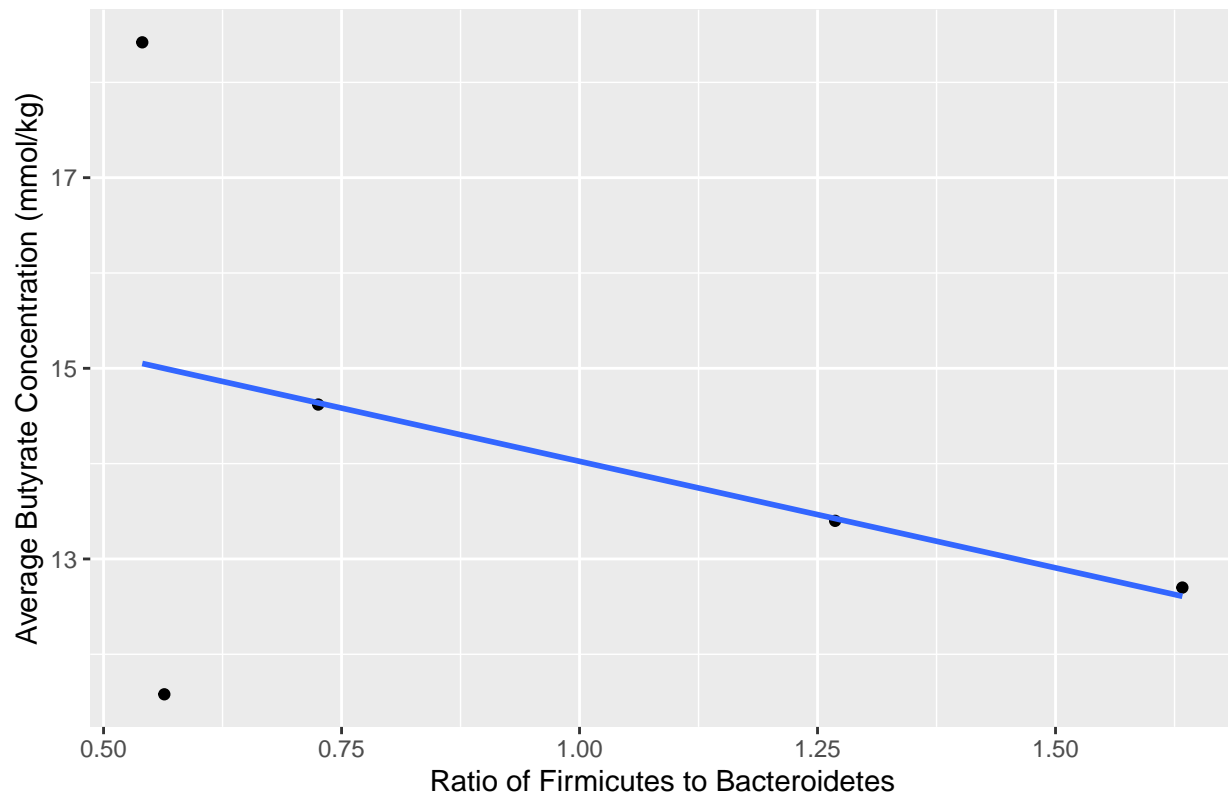

```
#TEST
obwk3fb_wk3but_test <- obese_phylum_scfa %>%
  lm(fb_ratio_wk3 ~ but_wk3, data = .) #test relationship
summary(obwk3fb_wk3but_test) #view results
```

```
## 
## Call:
## lm(formula = fb_ratio_wk3 ~ but_wk3, data = .)
## 
## Residuals:
##        1        2        3        4        5
## -0.18497 -0.08322  0.26603 -0.57585  0.57801
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.01291    1.38645   1.452    0.242
## but_wk3     -0.07541    0.09669  -0.780    0.492
## 
## Residual standard error: 0.5091 on 3 degrees of freedom
## Multiple R-squared:  0.1686, Adjusted R-squared:  -0.1086
## F-statistic: 0.6082 on 1 and 3 DF,  p-value: 0.4923
```

```
#p-value = 0.4923
#Adjusted R-squared = -0.1086
```

```
#Save Plot
save_plot(filename = "figures/obwk3fb_wk3but_plot.pdf", plot = obwk3fb_wk3but_plot, nrow = 1, ncol = 1,
```

The p-value for the linear correlation of the week 3 firmicutes to bacteroidetes ratio and the week 3 average butyrate concentration among "obese" participants was 0.4923, which is not statistically significant. Furthermore, the R value was an insignificant value close to zero (-0.1086). There does not appear to be a statistically significant linear relationship between the week 3 firmicutes to bacteroidetes ratio and the week 3 average butyrate concentration among participants whose BMI falls within the CDC's "Obese" category.

Wk3 F:B and Wk3 Propionate

```
#PLOT
obwk3fb_wk3pro_plot<-obese_phylum_scfa %>%
  ggplot(aes(x = fb_ratio_wk3,
             y = pro_wk3)) +
  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
 labs(title = "Week 3 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration",
         x = "Ratio of Firmicutes to Bacteroidetes",
         y = "Average Propionate Concentration (mmol/kg)")
print(obwk3fb_wk3pro_plot)
```

Week 3 Ratio of Firmicutes to Bacteroidetes by Propionate Concentration
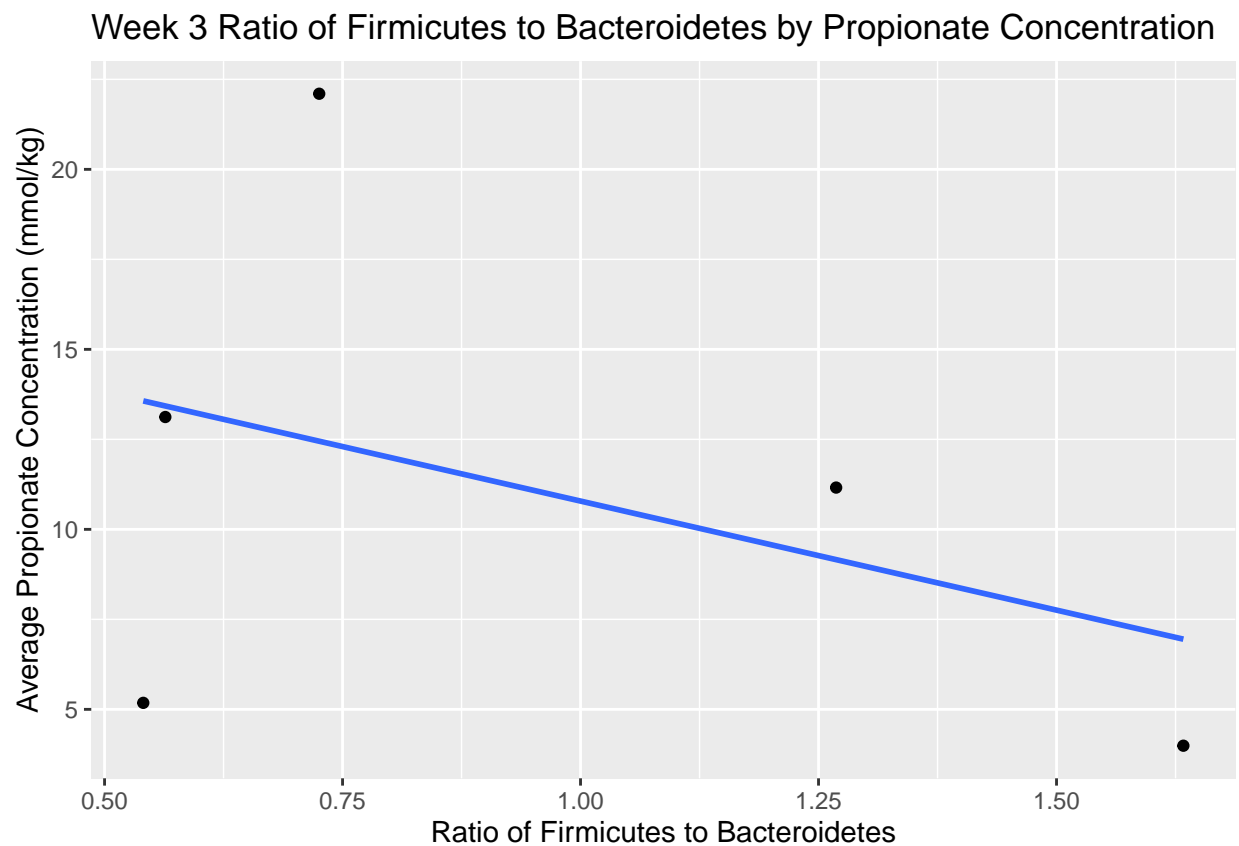

```
#TEST
obwk3fb_wk3pro_test <- obese_phylum_scfa %>%
```

```
  lm(fb_ratio_wk3 ~ pro_wk3, data = .) #test relationship
summary(obwk3fb_wk3pro_test) #view results
```

```
##
## Call:
## lm(formula = fb_ratio_wk3 ~ pro_wk3, data = .)
##
## Residuals:
##        1        2        3        4        5
##  0.07465 -0.56512  0.32348 -0.32846  0.49544
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.24509    0.45287   2.749   0.0708 .
## pro_wk3     -0.02689    0.03520  -0.764   0.5005
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5109 on 3 degrees of freedom
## Multiple R-squared:  0.1629, Adjusted R-squared:  -0.1162
## F-statistic: 0.5837 on 1 and 3 DF,  p-value: 0.5005
```

```
#p-value = 0.5005
#Adjusted R-squared = -0.1162

#Save Plot
save_plot(filename = "figures/obwk3fb_wk3pro_plot.pdf", plot = obwk3fb_wk3pro_plot, nrow = 1, ncol = 1,
```

The p-value for the linear correlation of the week 3 firmicutes to bacteroidetes ratio and the week 3 average propionate concentration among "obese" participants was 0.5005, which falls within a 95% confidence interval. However, the R value was an insignificant value close to zero (-0.1162). There does not appear to be a statistically significant linear relationship between the week 3 firmicutes to bacteroidetes ratio and the week 3 average propionate concentration among participants whose BMI falls within the CDC's "Obese" category.

#Final Conclusions Based on the linear correlation calculations and plots above, it does not appear that any of the above tested variables have a statistically significant relationship. It is possible that we failed to detect a significant relationship, because of the many limitations of this project. For one, the only type of relationship tested were linear correlations, but it is possible that these variables are related in other ways. Additionally, our analysis used weekly average measurements for scfa concentrations (mainly due to time contraints) rather than using individual scfa measurements. Finally, there was limited BMI data which narrowed our sample sizes quite a bit (participants were especially low in the underweight and obese BMI categories).

#Supplemental: Pre-Work This analysis was completed prior to analyzing BMI data and was not presented in our final presentation. These tests investigate potential relationships between butyrate and propionate and were eventaully subsetted by specific supplements. We also tried subsetting for only when butyrate increased and only when propionate decreased. We hypothesized that an increase in butyrate would be correlated with a decrease in propionate. #Broad ### Tidy Data #### Week 1 - Week 3 Change All

```
but_wide <- all_data_wkly %>%
  rename_all(tolower) %>%
  select(participant_id, semester, supplement_consumed, frequency, quantity_compliant, study_week, butyr
  filter(study_week == "week1" | study_week == "week3", quantity_compliant == "yes") %>% #keep only wee
```

```
    spread(study_week, butyrate_mean) %>%
    mutate(delta_butyrate = week1 - week3) #calculate difference in butyrate

pro_wide <- all_data_wkly %>%
    rename_all(tolower) %>%
    select(participant_id, semester, supplement_consumed, frequency, quantity_compliant, study_week, prop
    filter(study_week == "week1" | study_week == "week3", quantity_compliant == "yes") %>% #keep only wee
    spread(study_week, propionate_mean) %>%
    mutate(delta_propionate = week1 - week3) #calculate difference in propionate

delta_but_pro_wide <- inner_join(x = but_wide, y = pro_wide,
                    by = c("participant_id",
                           "semester", "supplement_consumed", "frequency")) %>%
    select(-starts_with("week"), -starts_with("quantity_compliant"))


incbut_broad_delta_wide <- delta_but_pro_wide %>%
    filter(delta_butyrate > 0) #subset for only increased burtyrate

decpro_broad_delta_wide <- delta_but_pro_wide %>%
    filter(delta_propionate < 0) #subset for only increased burtyrate
```

###Analysis ####Plot

```
linear_correlation_plot<-delta_but_pro_wide %>%
    ggplot(aes(x = delta_butyrate,
               y = delta_propionate)) +
    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("Delta Butyrate (mmol/kg)") +
    ylab("Delta Propionate (mmol/kg)")
print(linear_correlation_plot)
```

```
## Warning: Removed 130 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 130 rows containing missing values (geom_point).
```
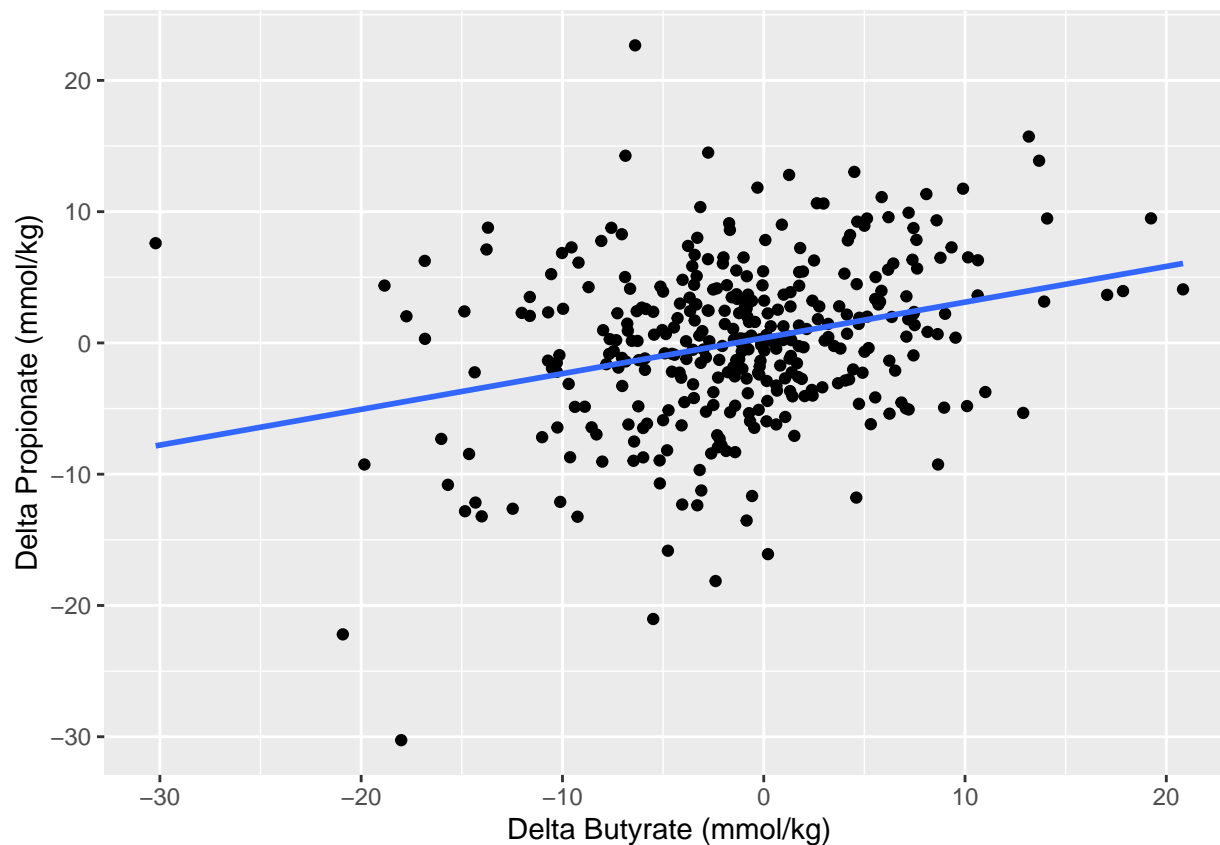
```
incbut_broad_plot<-incbut_broad_delta_wide %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  geom_point() + #puts data points to match x and y coordinates
  geom_smooth(method = "lm", #used to create a linear best fit line
              se = FALSE) + #hides confidence interval around line
  xlab("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(incbut_broad_plot)
```

Warning: Removed 17 rows containing non-finite values (stat_smooth).

Warning: Removed 17 rows containing missing values (geom_point).
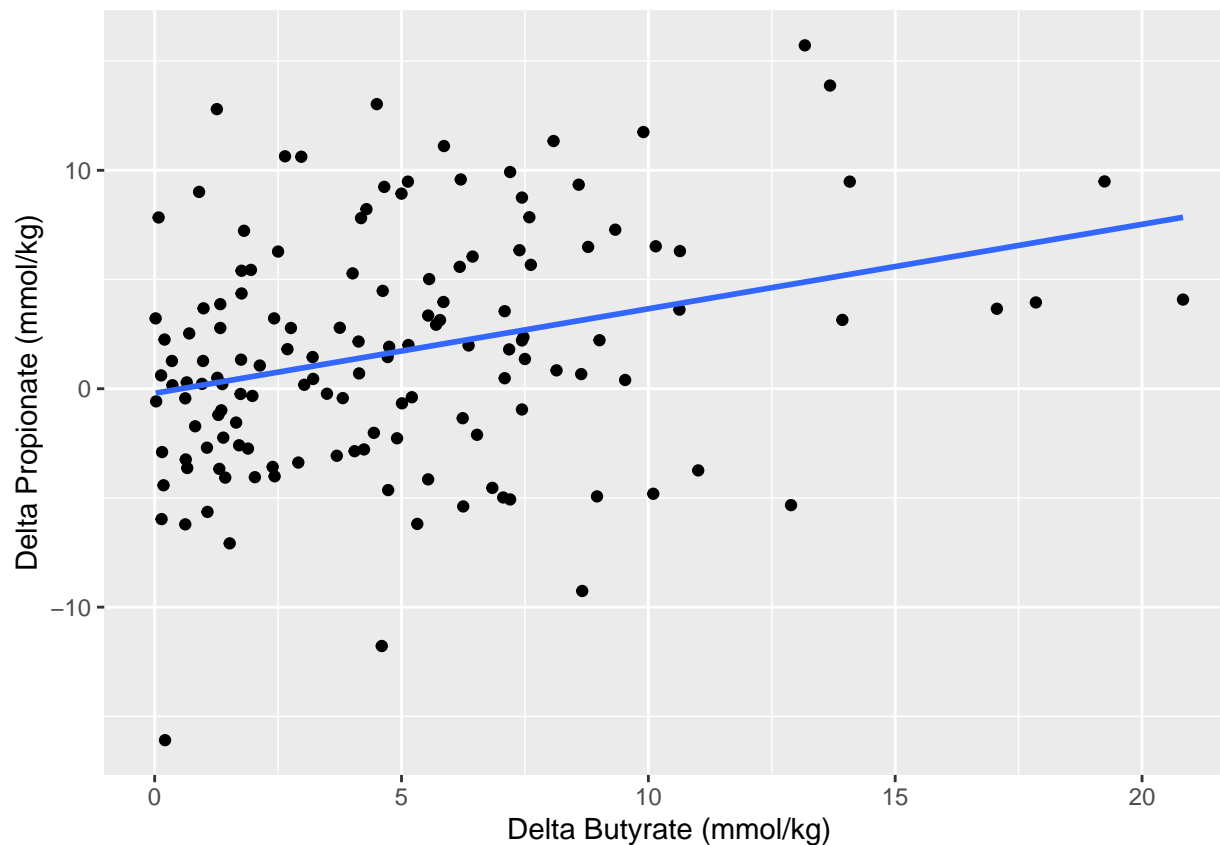
```
decpro_broad_plot<-decpro_broad_delta_wide %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  geom_point() + #puts data points to match x and y coordinates
  geom_smooth(method = "lm", #used to create a linear best fit line
              se = FALSE) + #hides confidence interval around line
  xlab("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(decpro_broad_plot)
```

Warning: Removed 2 rows containing non-finite values (stat_smooth).

Warning: Removed 2 rows containing missing values (geom_point).
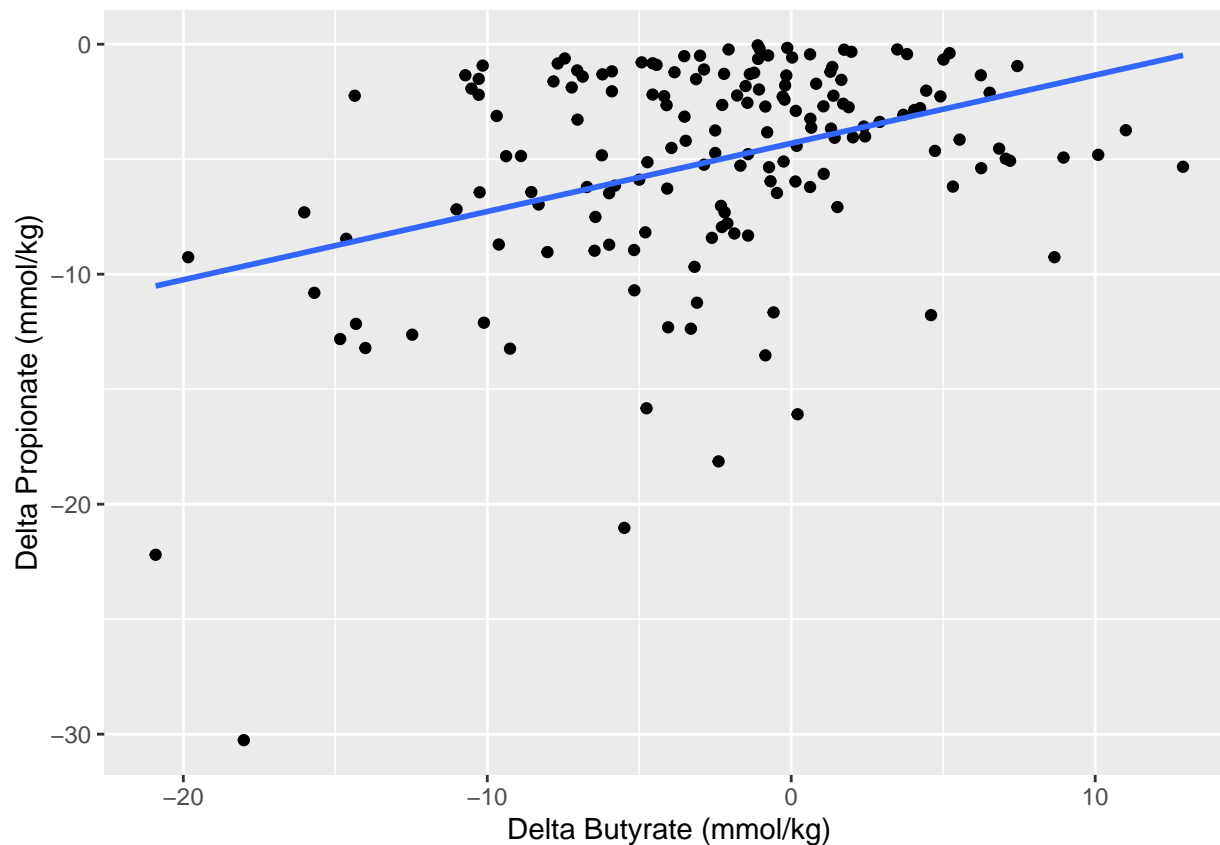
Assumptions

```r
nrow(delta_but_pro_wide) # sample size
```

```
## [1] 481
```

```r
shapiro.test(delta_but_pro_wide$delta_butyrate) #call column of df with values = vector
```

```
##
##  Shapiro-Wilk normality test
##
## data:  delta_but_pro_wide$delta_butyrate
## W = 0.98552, p-value = 0.0006516
```

```r
shapiro.test(delta_but_pro_wide$delta_propionate) #call column of df with values = vector
```

```
##
##  Shapiro-Wilk normality test
##
## data:  delta_but_pro_wide$delta_propionate
## W = 0.97482, p-value = 7.065e-06
```

```r
ggplot(delta_but_pro_wide, aes(x=delta_butyrate)) +
  geom_histogram() #histograms show the number of data points (count) at each value
```

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

```
## Warning: Removed 94 rows containing non-finite values (stat_bin).
```
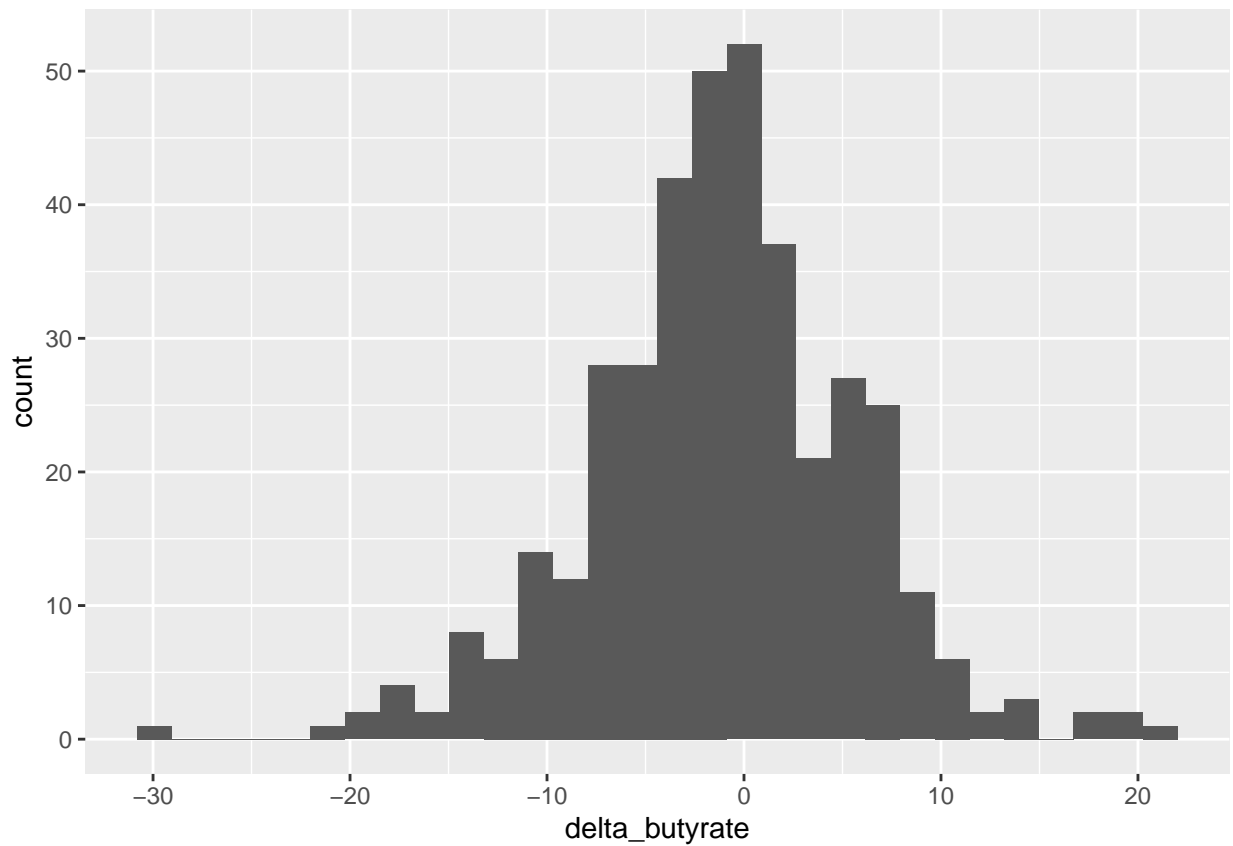


```
ggplot(delta_but_pro_wide, aes(x=delta_propionate)) +
  geom_histogram() #histograms show the number of data points (count) at each value
```

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

```
## Warning: Removed 124 rows containing non-finite values (stat_bin).
```
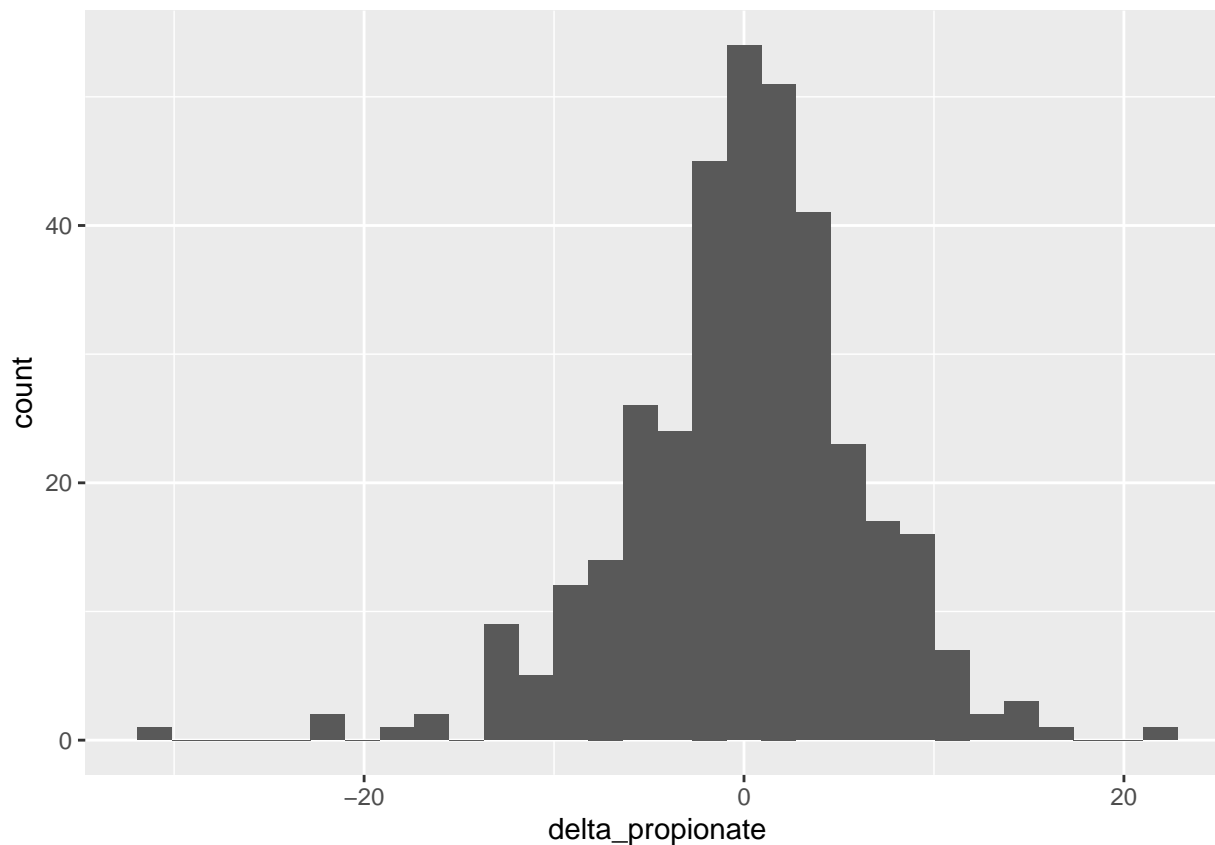
Test

```
all_data_lm <- delta_but_pro_wide %>%
  lm(delta_butyrate ~ delta_propionate, data = .) #test relationship
summary(all_data_lm)
```

```
##
## Call:
## lm(formula = delta_butyrate ~ delta_propionate, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -31.3254  -3.6260   0.3011   3.8289  20.8232
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.27816    0.34647  -3.689 0.000261 ***
## delta_propionate  0.31494    0.05505   5.721 2.28e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.491 on 349 degrees of freedom
##   (130 observations deleted due to missingness)
## Multiple R-squared:  0.08575,    Adjusted R-squared:  0.08313
## F-statistic: 32.73 on 1 and 349 DF,  p-value: 2.277e-08
```

```
incbut_broad_test <- incbut_broad_delta_wide %>%
  lm(delta_butyrate ~ delta_propionate, data = .) #test relationship
summary(incbut_broad_test) #view results
```

```
##
## Call:
## lm(formula = delta_butyrate ~ delta_propionate, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2408 -2.9809 -0.8744  2.1977 15.3592
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        4.5485     0.3534  12.871  < 2e-16 ***
## delta_propionate   0.2261     0.0624   3.623  0.00041 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.974 on 137 degrees of freedom
##   (17 observations deleted due to missingness)
## Multiple R-squared:  0.08742,    Adjusted R-squared:  0.08076
## F-statistic: 13.12 on 1 and 137 DF,  p-value: 0.0004098
```

```
#p-value = 0.01729
#Adjusted R-squared = 0.02205
```

```
decpro_broad_test <- decpro_broad_delta_wide %>%
  lm(delta_butyrate ~ delta_propionate, data = .) #test relationship
summary(decpro_broad_test) #view results
```

```
##
## Call:
## lm(formula = delta_butyrate ~ delta_propionate, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.2399  -3.4067   0.1421   3.1362  15.5741
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.08565    0.64929  -0.132    0.895
## delta_propionate  0.48752    0.09403   5.185 6.49e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.597 on 159 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.1446, Adjusted R-squared:  0.1392
## F-statistic: 26.88 on 1 and 159 DF,  p-value: 6.491e-07
```

```
#p-value = 0.05918
#Adjusted R-squared = 0.01359
```

#By Supplement ##BRMPS ###Tidy Data

```
BRMPS_delta_wide <- delta_but_pro_wide %>%
  filter(supplement_consumed == "BRMPS", frequency == "1xdaily") #subset for BRMPS 1x

inc_BRMPS_delta_wide <- delta_but_pro_wide %>%
  filter(supplement_consumed == "BRMPS", frequency == "1xdaily", delta_butyrate > 0) #subset for only i

dec_BRMPS_delta_wide <- delta_but_pro_wide %>%
  filter(supplement_consumed == "BRMPS", frequency == "1xdaily", delta_propionate < 0) #subset for only
```

###Analysis ####Assumptions

```
# check sample size
BRMPS_delta_wide %>%
  summarise(sample_size = n()) #use summarise() and n() to get counts of ppl
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1          66
```

```
#n=66
```

```
inc_BRMPS_delta_wide %>%
  summarise(sample_size = n()) #use summarise() and n() to get counts of ppl
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1          25
```

```
#n=39
```

```
dec_BRMPS_delta_wide %>%
  summarise(sample_size = n()) #use summarise() and n() to get counts of ppl
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1          26
```

```
#n=38
```

```
# normality
shapiro.test(BRMPS_delta_wide$delta_butyrate) #p-value = 0.5985
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  BRMPS_delta_wide$delta_butyrate
## W = 0.98444, p-value = 0.5985
```

```r
shapiro.test(BRMPS_delta_wide$delta_propionate) #p-value = 0.8763
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  BRMPS_delta_wide$delta_propionate
## W = 0.98977, p-value = 0.8763
```

```r
shapiro.test(inc_BRMPS_delta_wide$delta_butyrate) #p-value = 0.00022222
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  inc_BRMPS_delta_wide$delta_butyrate
## W = 0.88399, p-value = 0.008361
```

```r
shapiro.test(inc_BRMPS_delta_wide$delta_propionate) #p-value = 0.8143
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  inc_BRMPS_delta_wide$delta_propionate
## W = 0.9799, p-value = 0.8831
```

```r
shapiro.test(dec_BRMPS_delta_wide$delta_butyrate) #p-value = 0.7484
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  dec_BRMPS_delta_wide$delta_butyrate
## W = 0.97632, p-value = 0.7878
```

```r
shapiro.test(dec_BRMPS_delta_wide$delta_propionate) #p-value = 0.002383
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  dec_BRMPS_delta_wide$delta_propionate
## W = 0.91161, p-value = 0.0287
```

```r
# histogram
ggplot(BRMPS_delta_wide, aes(x=delta_butyrate)) +
  geom_histogram() #histogram for delta butyrate = very normal
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
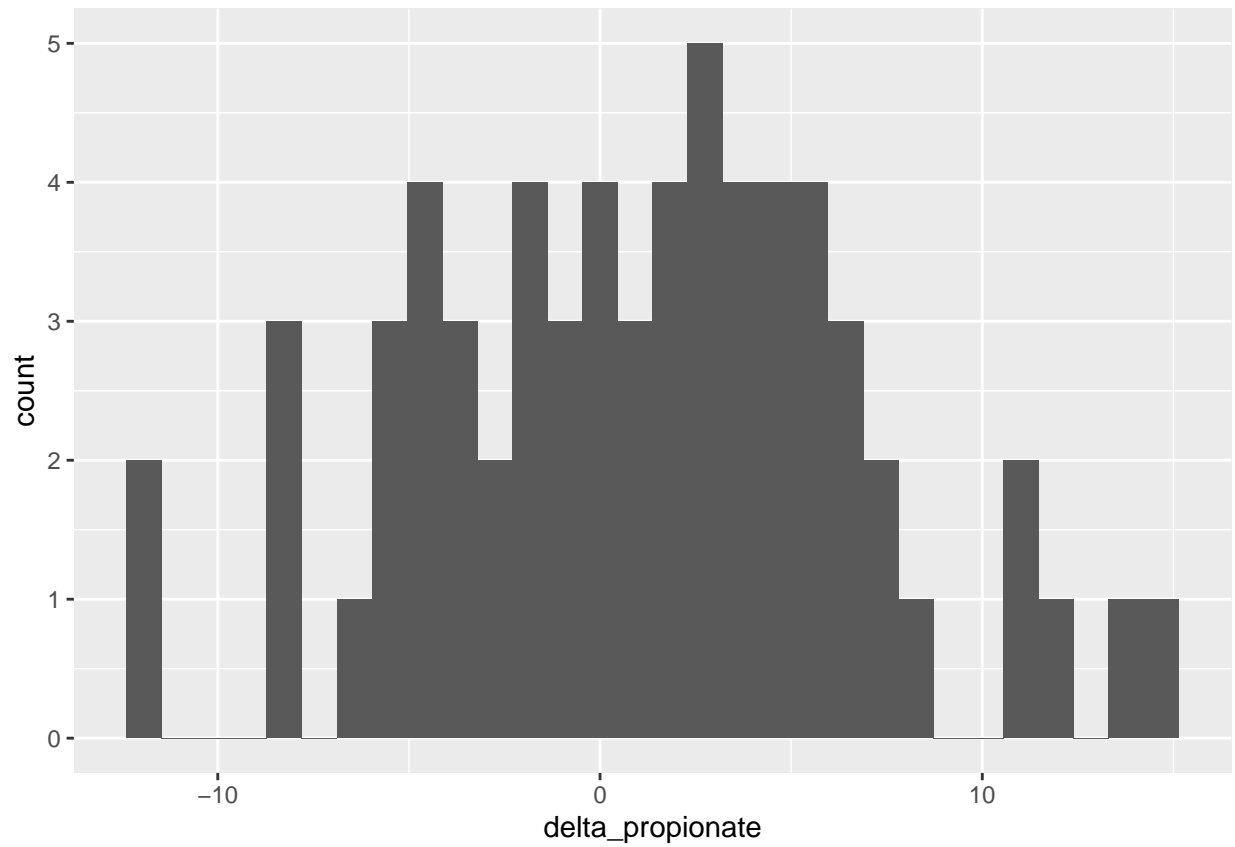
Warning: Removed 2 rows containing non-finite values (stat_bin).


```
ggplot(BRMPS_delta_wide, aes(x=delta_propionate)) +
  geom_histogram() #histogram for delta propionate = roughly normal
```
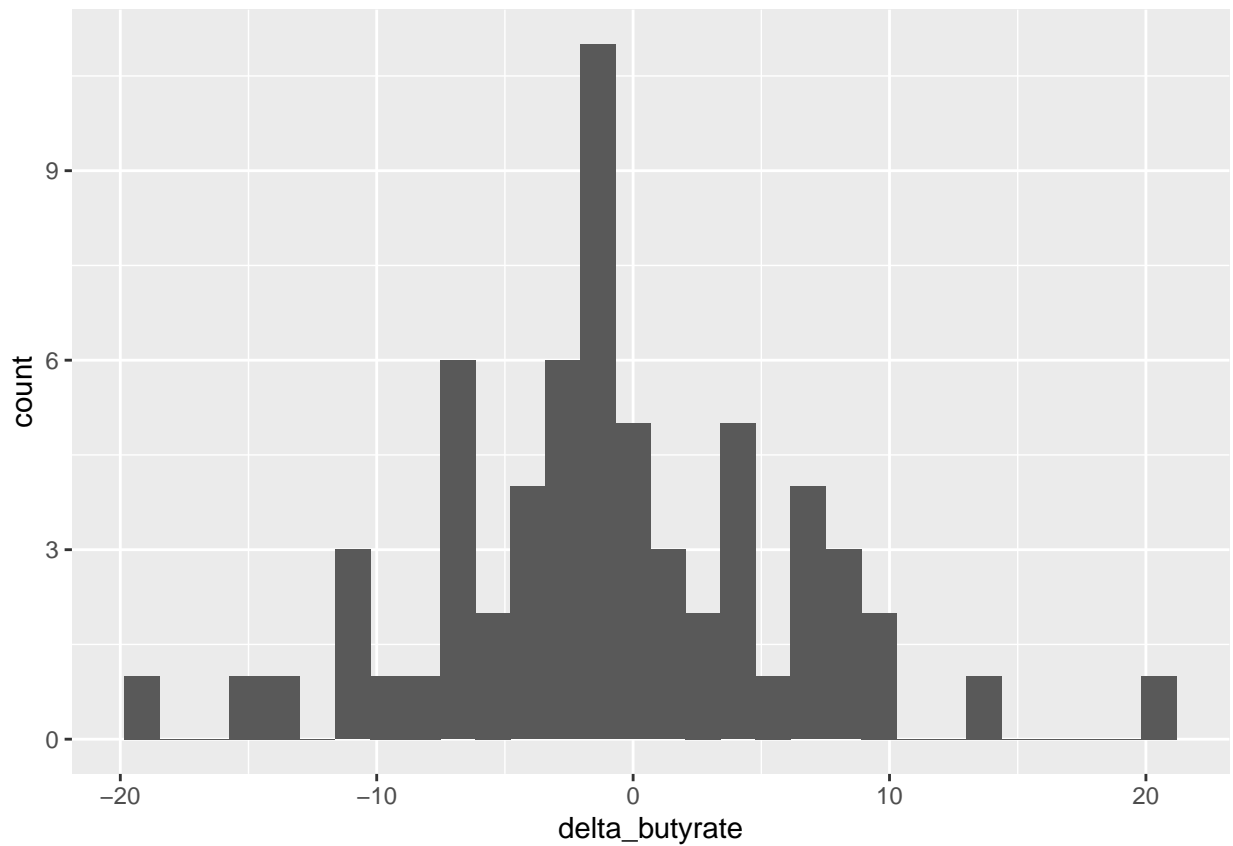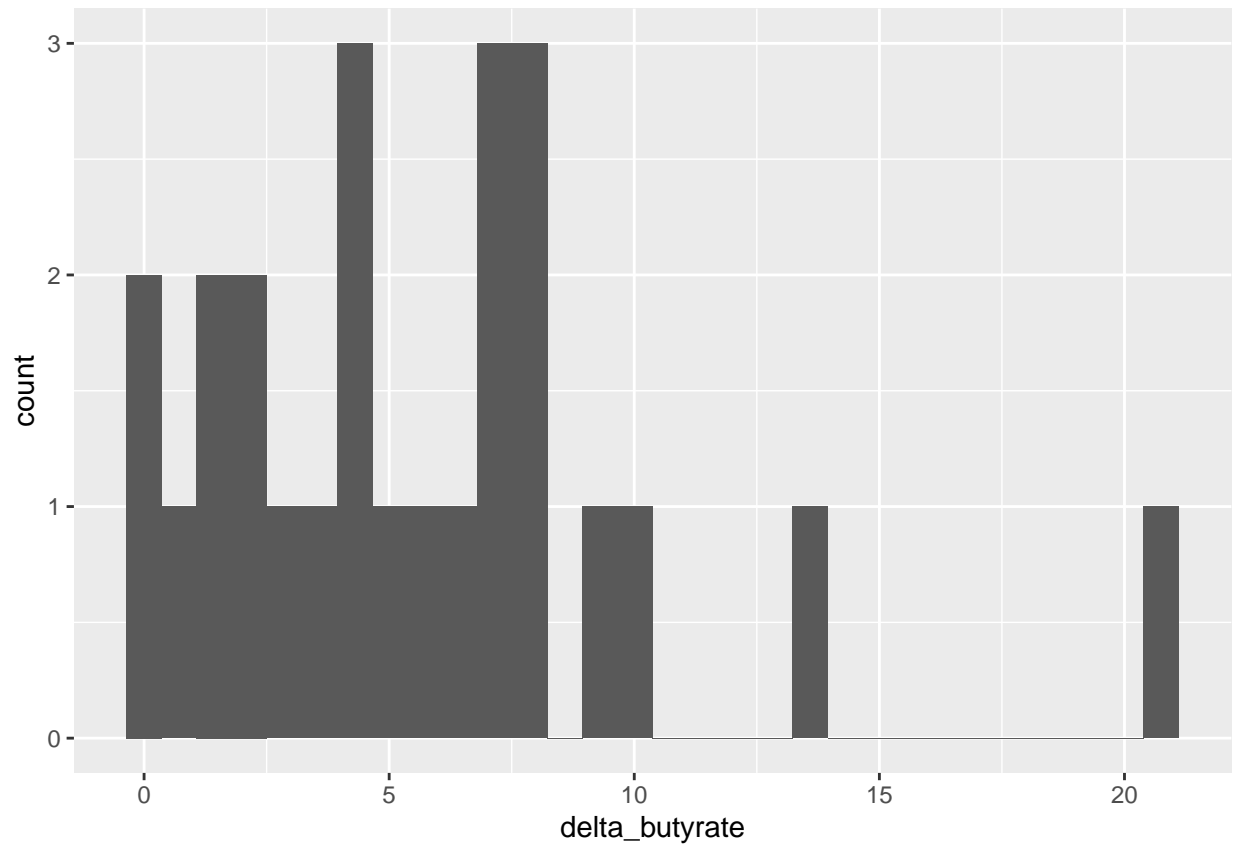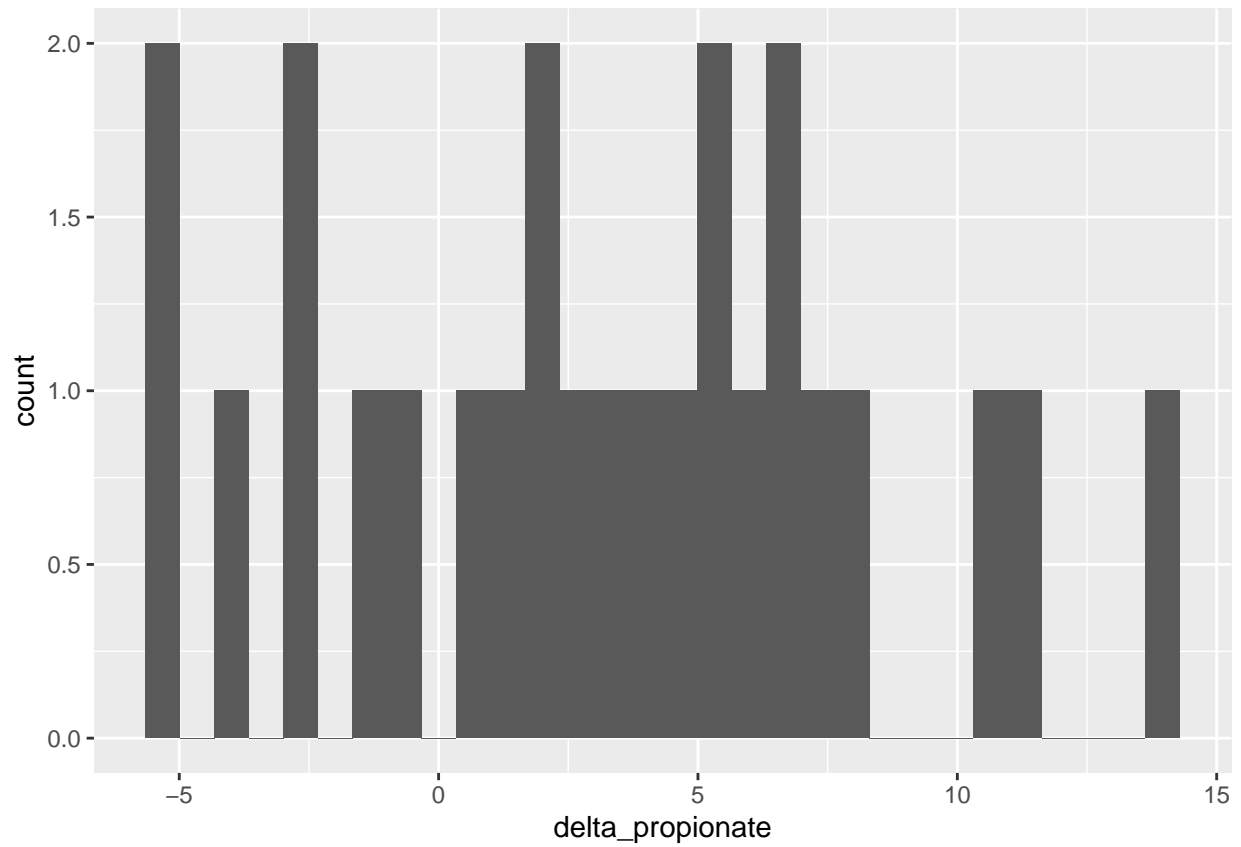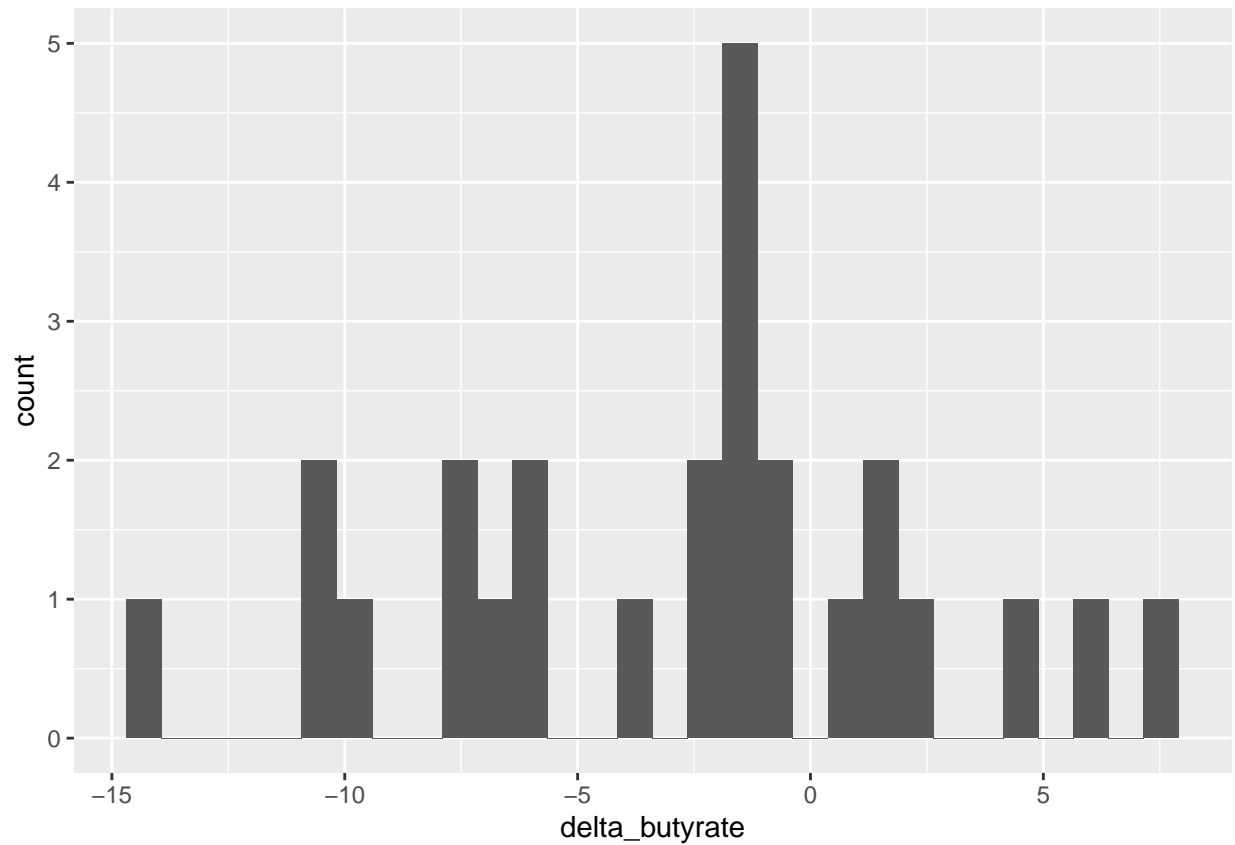
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

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

```
ggplot(inc_BRMPS_delta_wide, aes(x=delta_butyrate)) +
  geom_histogram() #histogram for delta butyrate = very skewed right
```

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

```
ggplot(inc_BRMPS_delta_wide, aes(x=delta_propionate)) +
  geom_histogram() #histogram for delta propionate = roughly normal
```

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

```r
ggplot(dec_BRMPS_delta_wide, aes(x=delta_butyrate)) +
  geom_histogram() #histogram for delta butyrate = roughly normal
```

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

```
ggplot(dec_BRMPS_delta_wide, aes(x=delta_propionate)) +
  geom_histogram() #histogram for delta propionate = skewed left
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
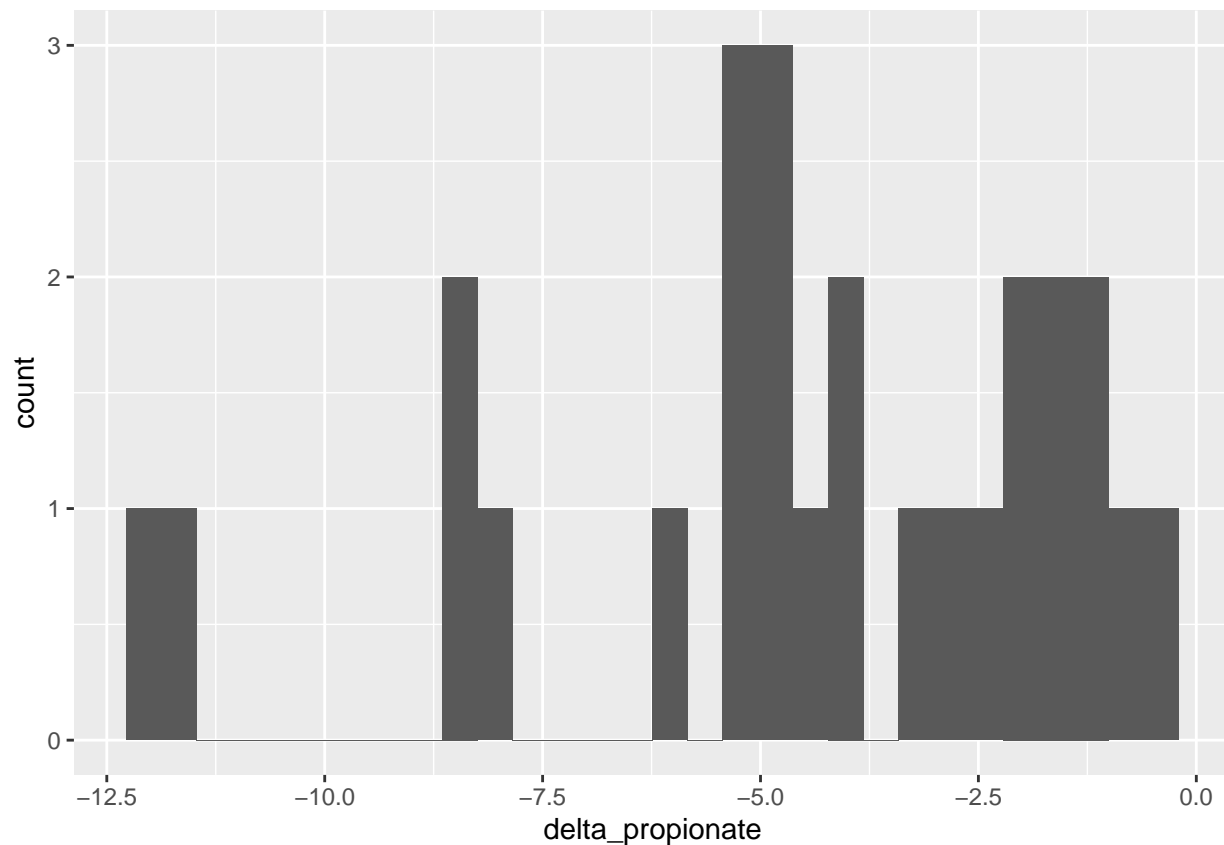
```r
# variances
var.test(x = BRMPS_delta_wide$delta_butyrate,
         y = BRMPS_delta_wide$delta_propionate,
         alternative = "two.sided") # p-value = 0.1289
```

```
## 
##  F test to compare two variances
## 
## data:  BRMPS_delta_wide$delta_butyrate and BRMPS_delta_wide$delta_propionate
## F = 1.47, num df = 63, denom df = 63, p-value = 0.1289
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.8930754 2.4196868
## sample estimates:
## ratio of variances
##            1.470021
```

```r
var.test(x = inc_BRMPS_delta_wide$delta_butyrate,
         y = inc_BRMPS_delta_wide$delta_propionate,
         alternative = "two.sided") # p-value = 0.1072
```

```
## 
##  F test to compare two variances
## 
## data:  inc_BRMPS_delta_wide$delta_butyrate and inc_BRMPS_delta_wide$delta_propionate
```

```
## F = 0.83061, num df = 24, denom df = 24, p-value = 0.653
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   0.366024 1.884884
## sample estimates:
## ratio of variances
##            0.83061
```

```r
var.test(x = dec_BRMPS_delta_wide$delta_butyrate,
         y = dec_BRMPS_delta_wide$delta_propionate,
         alternative = "two.sided") # p-value = 2.42e-05 UNEQUAL VARIANCES
```

```
##
##  F test to compare two variances
##
## data:  dec_BRMPS_delta_wide$delta_butyrate and dec_BRMPS_delta_wide$delta_propionate
## F = 2.9015, num df = 25, denom df = 25, p-value = 0.009921
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   1.300965 6.471320
## sample estimates:
## ratio of variances
##           2.901544
```
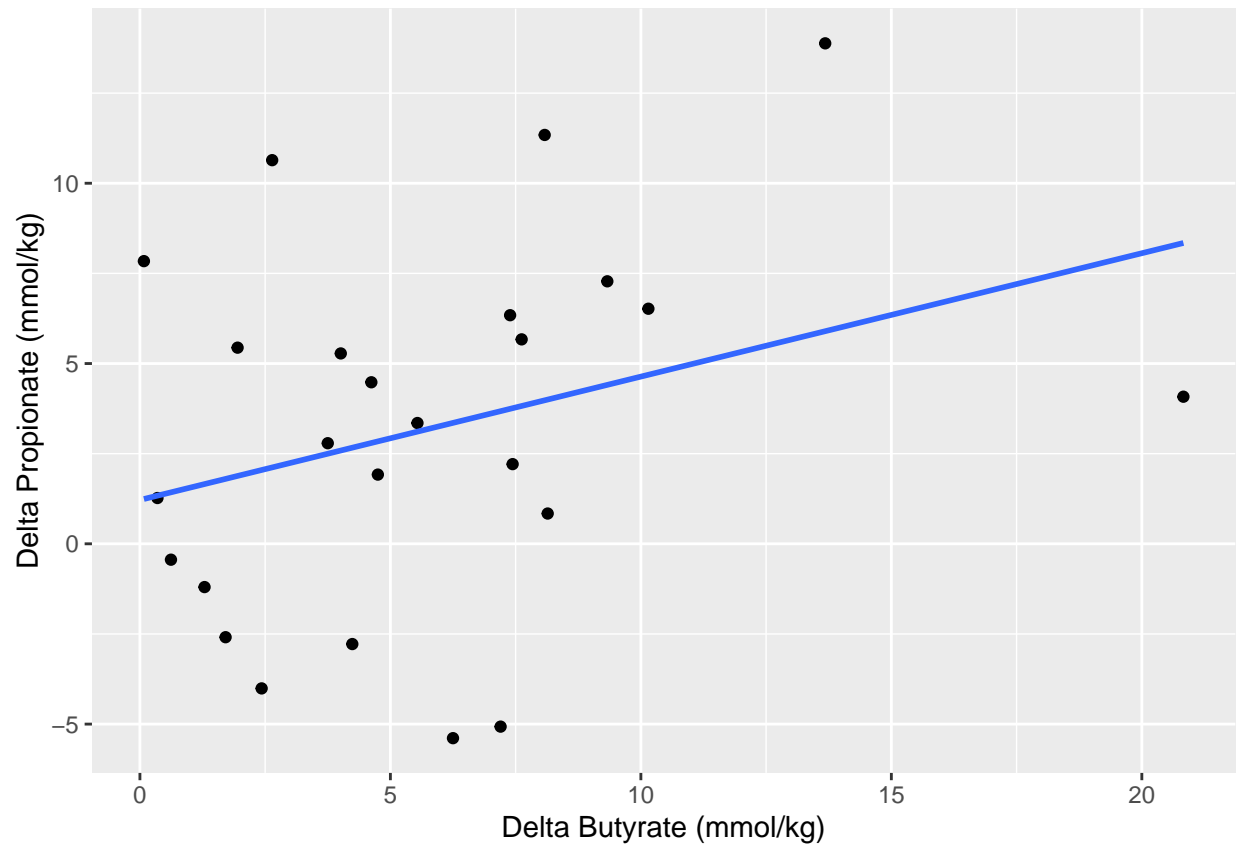
Plot

```r
BRMPS_plot<-BRMPS_delta_wide %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  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("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(BRMPS_plot)
```

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

```
## Warning: Removed 2 rows containing missing values (geom_point).
```
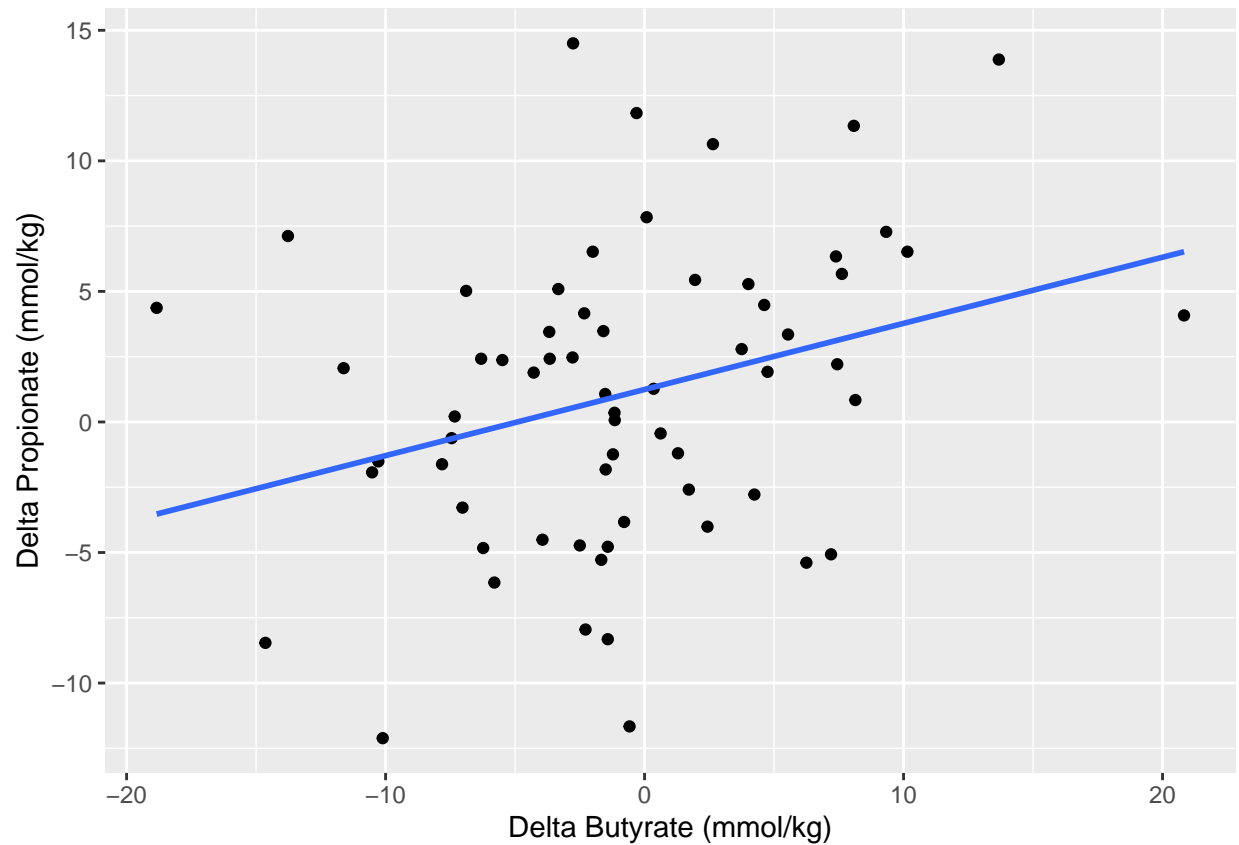
```
inc_BRMPS_plot<-inc_BRMPS_delta_wide %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  geom_point() + #puts data points to match x and y coordinates
  geom_smooth(method = "lm", #used to create a linear best fit line
              se = FALSE) + #hides confidence interval around line
  xlab("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(inc_BRMPS_plot)
```

```
dec_BRMPS_plot<-dec_BRMPS_delta_wide %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  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("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(dec_BRMPS_plot)
```
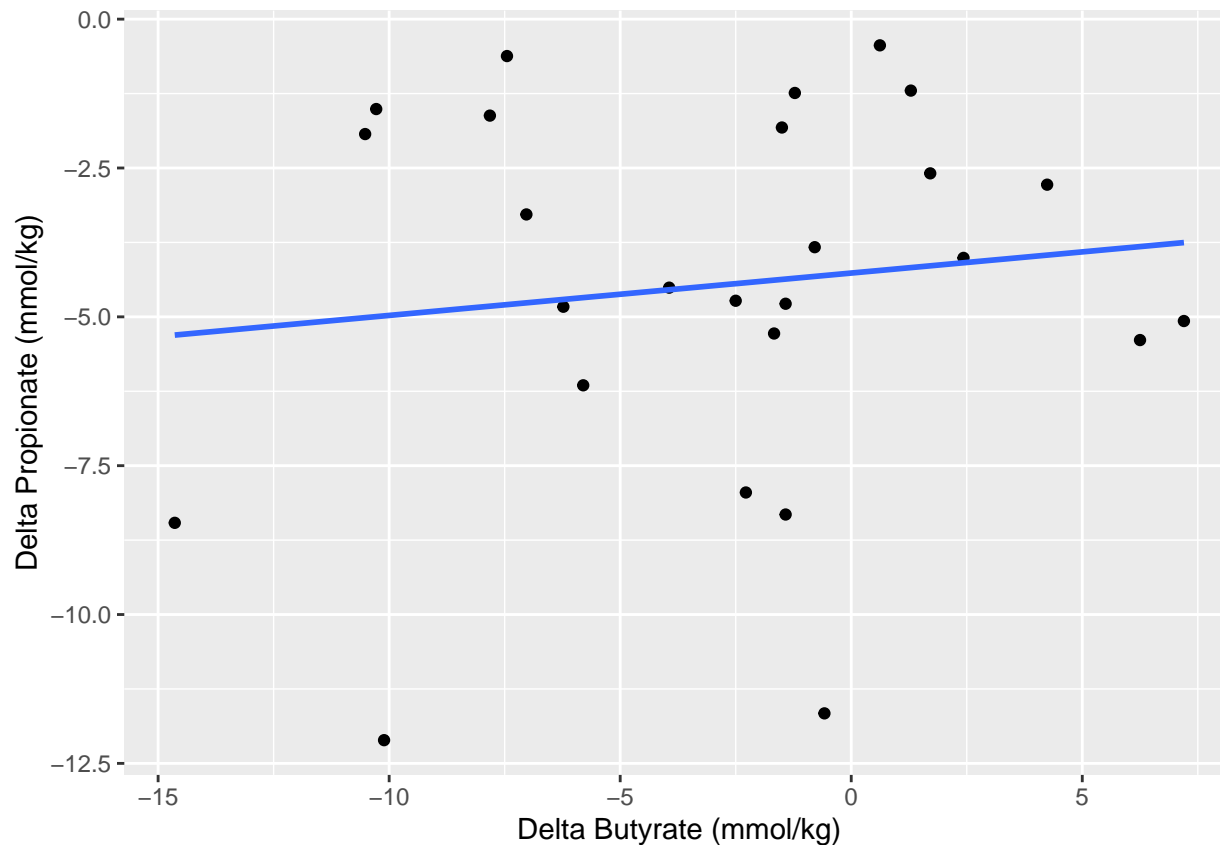
Test

```
BRMPS_test <- BRMPS_delta_wide %>%
  lm(delta_butyrate ~ delta_propionate, data = .) #test relationship
summary(BRMPS_test) #view results
```

```
##
## Call:
## lm(formula = delta_butyrate ~ delta_propionate, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.2410  -3.8181   0.2266   4.3802  20.5369
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -1.2259     0.8479  -1.446   0.1533
## delta_propionate   0.3723     0.1465   2.540   0.0136 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.675 on 62 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.09428,    Adjusted R-squared:  0.07967
## F-statistic: 6.454 on 1 and 62 DF,  p-value: 0.01359
```

```
#p-value = 0.01359
#Adjusted R-squared = 0.07967

inc_BRMPS_test <- inc_BRMPS_delta_wide %>%
  lm(delta_butyrate ~ delta_propionate, data = .) #test relationship
summary(inc_BRMPS_test) #view results
```

```
##
## Call:
## lm(formula = delta_butyrate ~ delta_propionate, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0060 -2.4114 -0.2698  2.4032 14.8128
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        4.8576     1.0681   4.548 0.000144 ***
## delta_propionate   0.2842     0.1806   1.574 0.129098
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.499 on 23 degrees of freedom
## Multiple R-squared:  0.09726,    Adjusted R-squared:  0.05801
## F-statistic: 2.478 on 1 and 23 DF,  p-value: 0.1291
```

```
#p-value = 0.9439
#Adjusted R-squared = -0.02689

dec_BRMPS_test <- dec_BRMPS_delta_wide %>%
  lm(delta_butyrate ~ delta_propionate, data = .) #test relationship
summary(dec_BRMPS_test) #view results
```

```
##
## Call:
## lm(formula = delta_butyrate ~ delta_propionate, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.990  -4.169   1.102   3.235  10.150
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -1.9037     1.8751  -1.015    0.320
## delta_propionate   0.2064     0.3451   0.598    0.555
##
## Residual standard error: 5.445 on 24 degrees of freedom
## Multiple R-squared:  0.01468,    Adjusted R-squared:  -0.02638
## F-statistic: 0.3576 on 1 and 24 DF,  p-value: 0.5555
```

```
#p-value = 0.1572
#Adjusted R-squared = 0.02853
```

LOODAT ### Tidy Data

```
but_wide_LOODAT <- all_data_wkly %>%
  rename_all(tolower) %>%
  select(participant_id, semester, supplement_consumed, frequency, quantity_compliant, study_week, buty
  filter(study_week == "week1" | study_week == "week3", quantity_compliant == "yes", frequency == "1xda
  spread(study_week, butyrate_mean) %>%
  mutate(delta_butyrate = week3 - week1) #calculate difference in butyrate
but_wide_LOODAT
```

```
## # A tibble: 16 x 8
##    participant_id semester supplement_cons~ frequency quantity_compli~
##    <chr>          <chr>    <chr>            <chr>     <chr>
##  1 U700           Winter2~ LOODAT           1xdaily   yes
##  2 U701           Winter2~ LOODAT           1xdaily   yes
##  3 U706           Winter2~ LOODAT           1xdaily   yes
##  4 U707           Winter2~ LOODAT           1xdaily   yes
##  5 U708           Winter2~ LOODAT           1xdaily   yes
##  6 U710           Winter2~ LOODAT           1xdaily   yes
##  7 U711           Winter2~ LOODAT           1xdaily   yes
##  8 U712           Winter2~ LOODAT           1xdaily   yes
##  9 U713           Winter2~ LOODAT           1xdaily   yes
## 10 U715           Winter2~ LOODAT           1xdaily   yes
## 11 U716           Winter2~ LOODAT           1xdaily   yes
## 12 U717           Winter2~ LOODAT           1xdaily   yes
## 13 U719           Winter2~ LOODAT           1xdaily   yes
## 14 U720           Winter2~ LOODAT           1xdaily   yes
## 15 U721           Winter2~ LOODAT           1xdaily   yes
## 16 U723           Winter2~ LOODAT           1xdaily   yes
## # ... with 3 more variables: week1 <dbl>, week3 <dbl>,
## #   delta_butyrate <dbl>
```

```
pro_wide_LOODAT <- all_data_wkly %>%
  rename_all(tolower) %>%
  select(participant_id, semester, supplement_consumed, frequency, quantity_compliant, study_week, prop
  filter(study_week == "week1" | study_week == "week3", quantity_compliant == "yes", frequency == "1xda
  spread(study_week, propionate_mean) %>%
  mutate(delta_propionate = week3 - week1) #calculate difference in propionate
pro_wide_LOODAT
```

```
## # A tibble: 16 x 8
##    participant_id semester supplement_cons~ frequency quantity_compli~
##    <chr>          <chr>    <chr>            <chr>     <chr>
##  1 U700           Winter2~ LOODAT           1xdaily   yes
##  2 U701           Winter2~ LOODAT           1xdaily   yes
##  3 U706           Winter2~ LOODAT           1xdaily   yes
##  4 U707           Winter2~ LOODAT           1xdaily   yes
##  5 U708           Winter2~ LOODAT           1xdaily   yes
##  6 U710           Winter2~ LOODAT           1xdaily   yes
##  7 U711           Winter2~ LOODAT           1xdaily   yes
##  8 U712           Winter2~ LOODAT           1xdaily   yes
##  9 U713           Winter2~ LOODAT           1xdaily   yes
## 10 U715           Winter2~ LOODAT           1xdaily   yes
```

```
## 11 U716           Winter2~ LOODAT         1xdaily   yes
## 12 U717           Winter2~ LOODAT         1xdaily   yes
## 13 U719           Winter2~ LOODAT         1xdaily   yes
## 14 U720           Winter2~ LOODAT         1xdaily   yes
## 15 U721           Winter2~ LOODAT         1xdaily   yes
## 16 U723           Winter2~ LOODAT         1xdaily   yes
## # ... with 3 more variables: week1 <dbl>, week3 <dbl>,
## #   delta_propionate <dbl>
```

```r
delta_but_pro_wide_LOODAT <- inner_join(x = but_wide_LOODAT, y = pro_wide_LOODAT,
                   by = c("participant_id",
                          "semester", "supplement_consumed")) %>%
  select(-starts_with("week"))
delta_but_pro_wide_LOODAT
```

```
## # A tibble: 16 x 9
##    participant_id semester supplement_cons~ frequency.x quantity_compli~
##    <chr>          <chr>    <chr>            <chr>       <chr>
##  1 U700           Winter2~ LOODAT           1xdaily     yes
##  2 U701           Winter2~ LOODAT           1xdaily     yes
##  3 U706           Winter2~ LOODAT           1xdaily     yes
##  4 U707           Winter2~ LOODAT           1xdaily     yes
##  5 U708           Winter2~ LOODAT           1xdaily     yes
##  6 U710           Winter2~ LOODAT           1xdaily     yes
##  7 U711           Winter2~ LOODAT           1xdaily     yes
##  8 U712           Winter2~ LOODAT           1xdaily     yes
##  9 U713           Winter2~ LOODAT           1xdaily     yes
## 10 U715           Winter2~ LOODAT           1xdaily     yes
## 11 U716           Winter2~ LOODAT           1xdaily     yes
## 12 U717           Winter2~ LOODAT           1xdaily     yes
## 13 U719           Winter2~ LOODAT           1xdaily     yes
## 14 U720           Winter2~ LOODAT           1xdaily     yes
## 15 U721           Winter2~ LOODAT           1xdaily     yes
## 16 U723           Winter2~ LOODAT           1xdaily     yes
## # ... with 4 more variables: delta_butyrate <dbl>, frequency.y <chr>,
## #   quantity_compliant.y <chr>, delta_propionate <dbl>
```

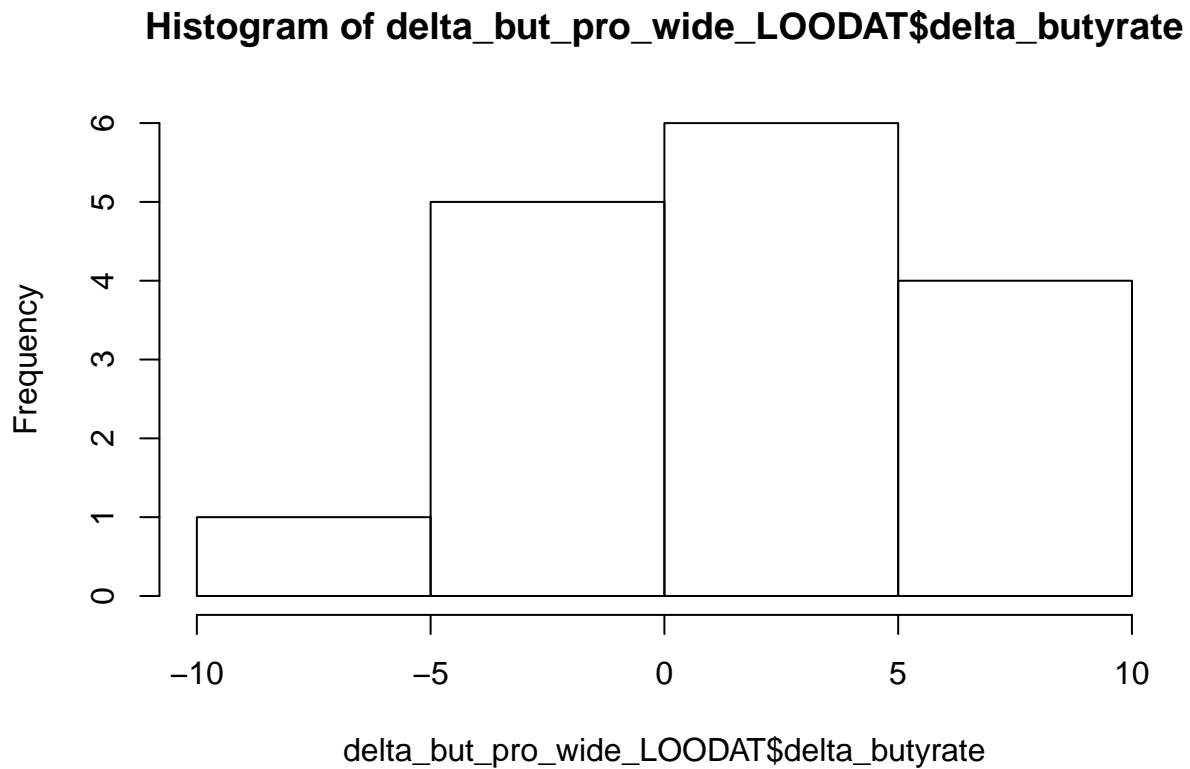### Analysis #### Assumptions

```r
#Sample Size
delta_but_pro_wide_LOODAT %>%
  summarise(sample_size = n()) #n=16
```

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

```r
#Normality
#Butyrate:
shapiro.test(delta_but_pro_wide_LOODAT$delta_butyrate)
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  delta_but_pro_wide_LOODAT$delta_butyrate
## W = 0.9732, p-value = 0.8874
```
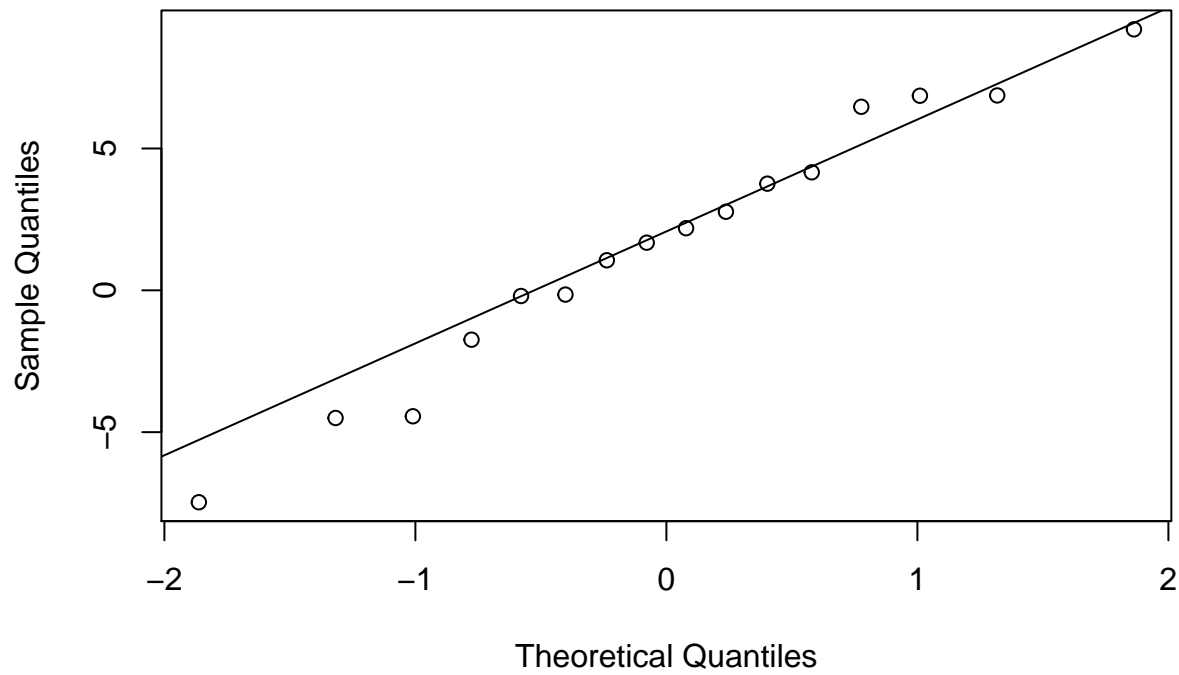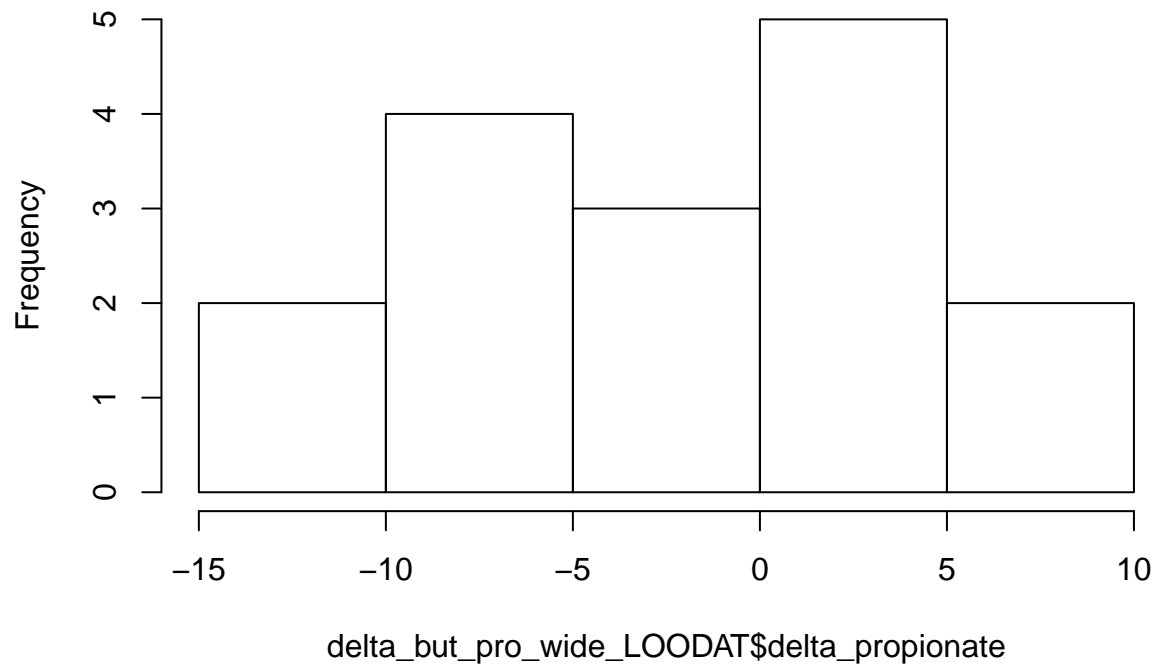
hist(delta_but_pro_wide_LOODAT**$**delta_butyrate)

Histogram of delta_but_pro_wide_LOODAT$delta_butyrate

qqnorm(delta_but_pro_wide_LOODAT**$**delta_butyrate)
qqline(delta_but_pro_wide_LOODAT**$**delta_butyrate) *#p=0.8874. Normality supported.*

Normal Q–Q Plot


```
#Propionate:
shapiro.test(delta_but_pro_wide_LOODAT$delta_propionate)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  delta_but_pro_wide_LOODAT$delta_propionate
## W = 0.97347, p-value = 0.8912
```
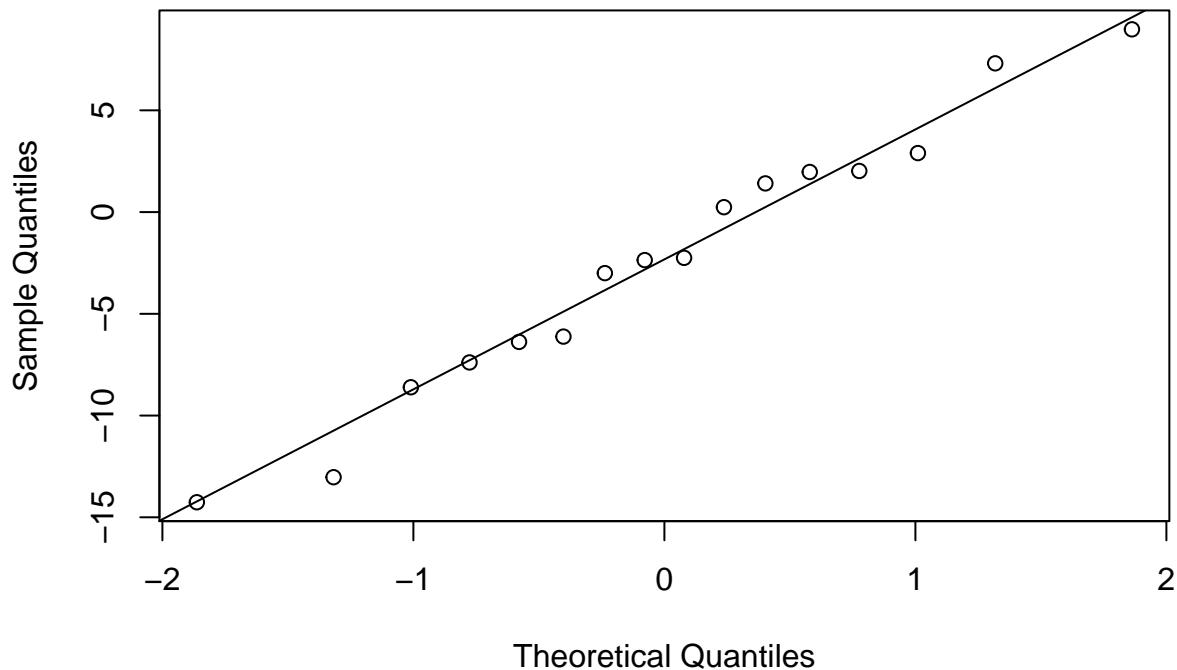
```
hist(delta_but_pro_wide_LOODAT$delta_propionate)
```

Histogram of delta_but_pro_wide_LOODAT$delta_propionate

delta_but_pro_wide_LOODAT$delta_propionate

```
qqnorm(delta_but_pro_wide_LOODAT$delta_propionate)
qqline(delta_but_pro_wide_LOODAT$delta_propionate) #p=0.8912. Normality supported
```

Normal Q–Q Plot


```
#Equality of Variances
var.test(delta_but_pro_wide_LOODAT$delta_butyrate, delta_but_pro_wide_LOODAT$delta_propionate, alternat
```

```
##
##  F test to compare two variances
##
## data:  delta_but_pro_wide_LOODAT$delta_butyrate and delta_but_pro_wide_LOODAT$delta_propionate
## F = 0.49078, num df = 15, denom df = 15, p-value = 0.1796
## alternative hypothesis: true ratio of variances is not equal to 1
## 90 percent confidence interval:
##  0.2041992 1.1795684
## sample estimates:
## ratio of variances
##           0.490782
```

```
var.test(delta_but_pro_wide_LOODAT$delta_butyrate, delta_but_pro_wide_LOODAT$delta_propionate, alternat
```

```
##
##  F test to compare two variances
##
## data:  delta_but_pro_wide_LOODAT$delta_butyrate and delta_but_pro_wide_LOODAT$delta_propionate
## F = 0.49078, num df = 15, denom df = 15, p-value = 0.9102
## alternative hypothesis: true ratio of variances is greater than 1
## 90 percent confidence interval:
##  0.248848       Inf
```

```
## sample estimates:
## ratio of variances
##            0.490782
```

```r
var.test(delta_but_pro_wide_LOODAT$delta_butyrate, delta_but_pro_wide_LOODAT$delta_propionate, alternat
```

```
##
##  F test to compare two variances
##
## data:  delta_but_pro_wide_LOODAT$delta_butyrate and delta_but_pro_wide_LOODAT$delta_propionate
## F = 0.49078, num df = 15, denom df = 15, p-value = 0.08982
## alternative hypothesis: true ratio of variances is less than 1
## 90 percent confidence interval:
##  0.0000000 0.9679278
## sample estimates:
## ratio of variances
##            0.490782
```

```r
#-------------------------------------------------------------------------------
#Assumptions selecting for butyrate increase:
delta_but_pro_wide_LOODAT_dbup <- delta_but_pro_wide_LOODAT %>%
  filter(delta_butyrate > 0) %>%
  summarise(sample_size = n()) #n=10
delta_but_pro_wide_LOODAT_dbup
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1          10
```

```r
delta_but_pro_wide_LOODAT_dbup <- delta_but_pro_wide_LOODAT %>%
  filter(delta_butyrate > 0)

#Normality
#Butyrate:
shapiro.test(delta_but_pro_wide_LOODAT_dbup$delta_butyrate)
```
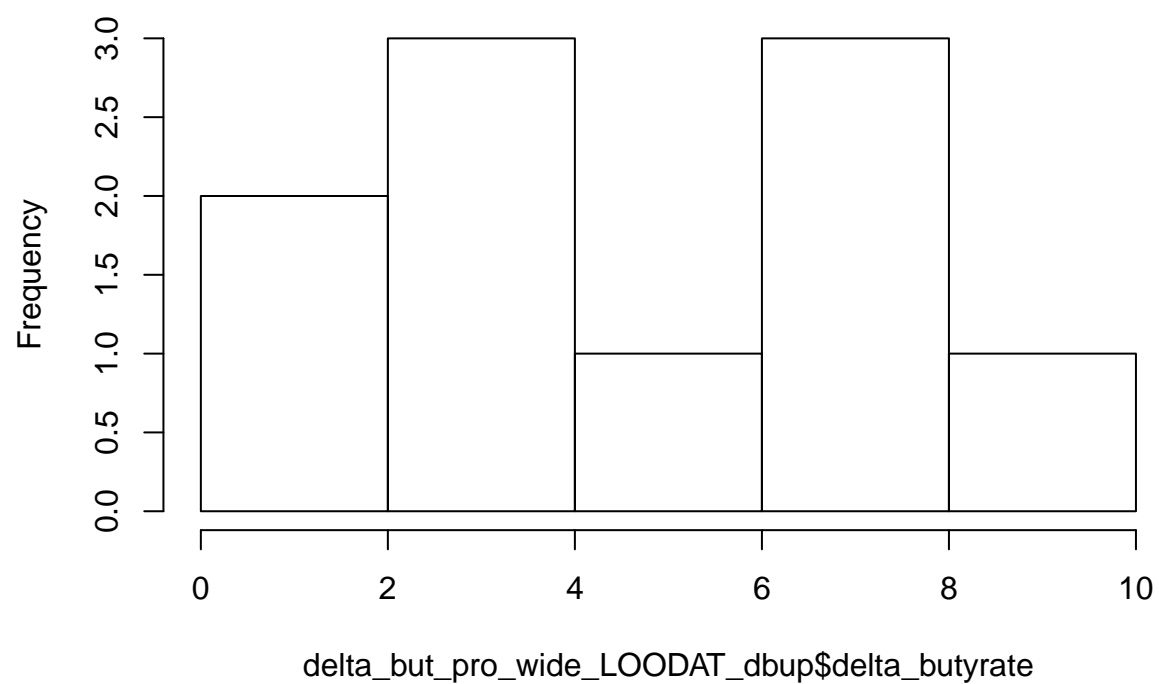
```
##
##  Shapiro-Wilk normality test
##
## data:  delta_but_pro_wide_LOODAT_dbup$delta_butyrate
## W = 0.93429, p-value = 0.4914
```

```r
hist(delta_but_pro_wide_LOODAT_dbup$delta_butyrate)
```
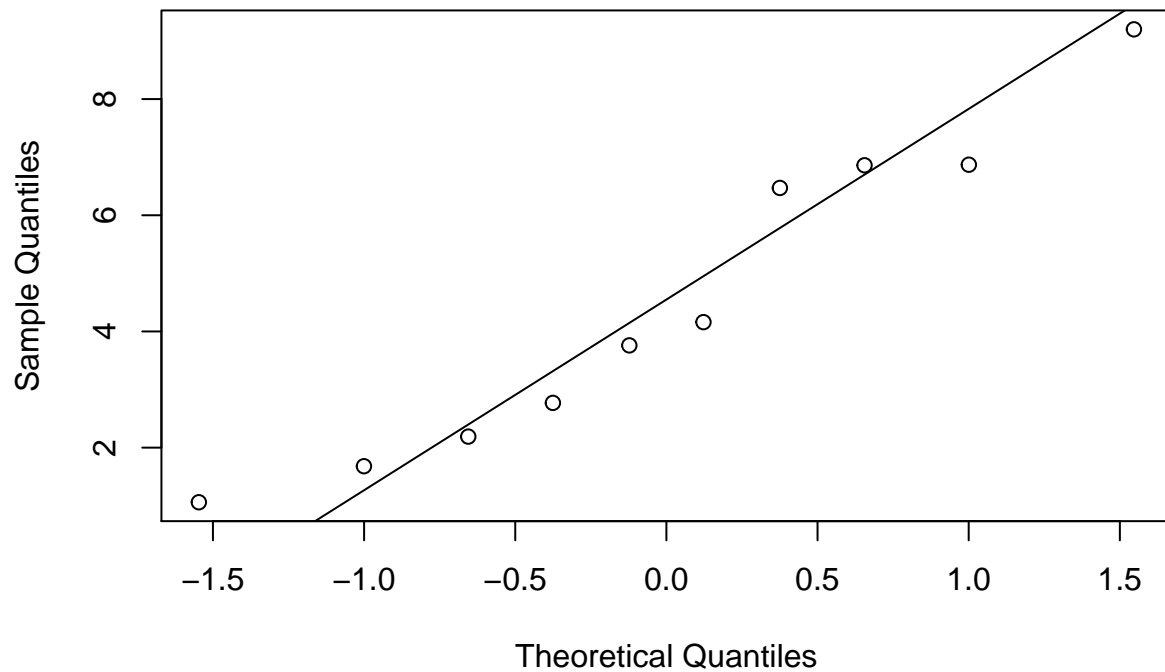
Histogram of delta_but_pro_wide_LOODAT_dbup$delta_butyrate

delta_but_pro_wide_LOODAT_dbup$delta_butyrate

```
qqnorm(delta_but_pro_wide_LOODAT_dbup$delta_butyrate)
qqline(delta_but_pro_wide_LOODAT_dbup$delta_butyrate) #p=0.4914. Normality supported.
```
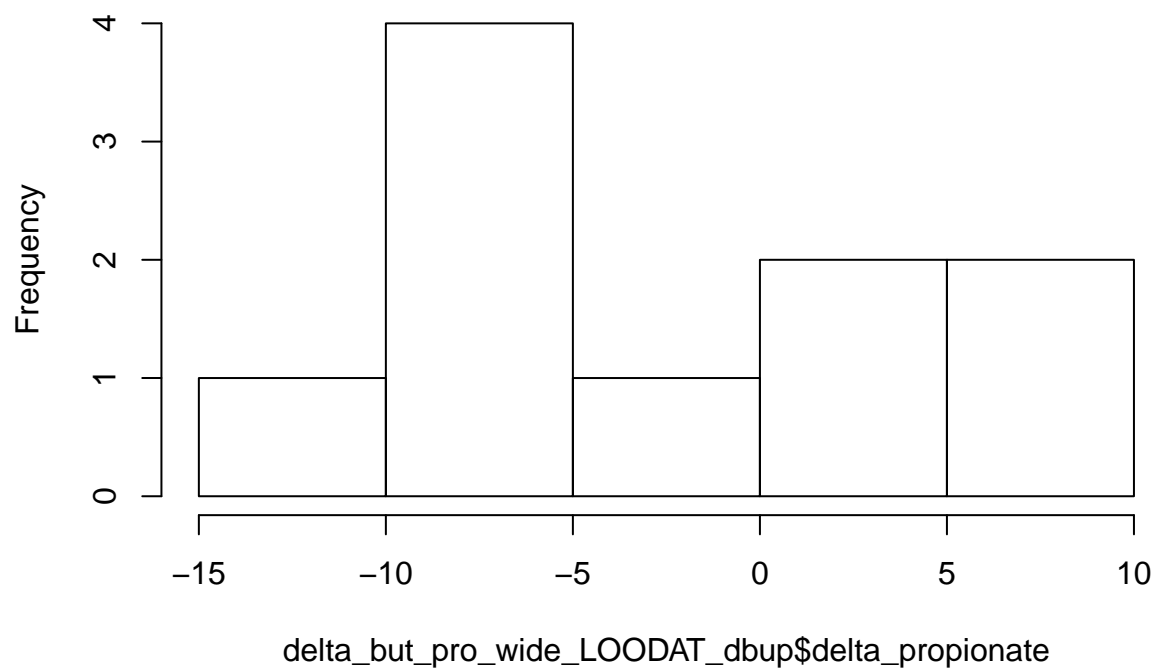
Normal Q–Q Plot


```
#Propionate:
shapiro.test(delta_but_pro_wide_LOODAT_dbup$delta_propionate)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  delta_but_pro_wide_LOODAT_dbup$delta_propionate
## W = 0.95511, p-value = 0.729
```
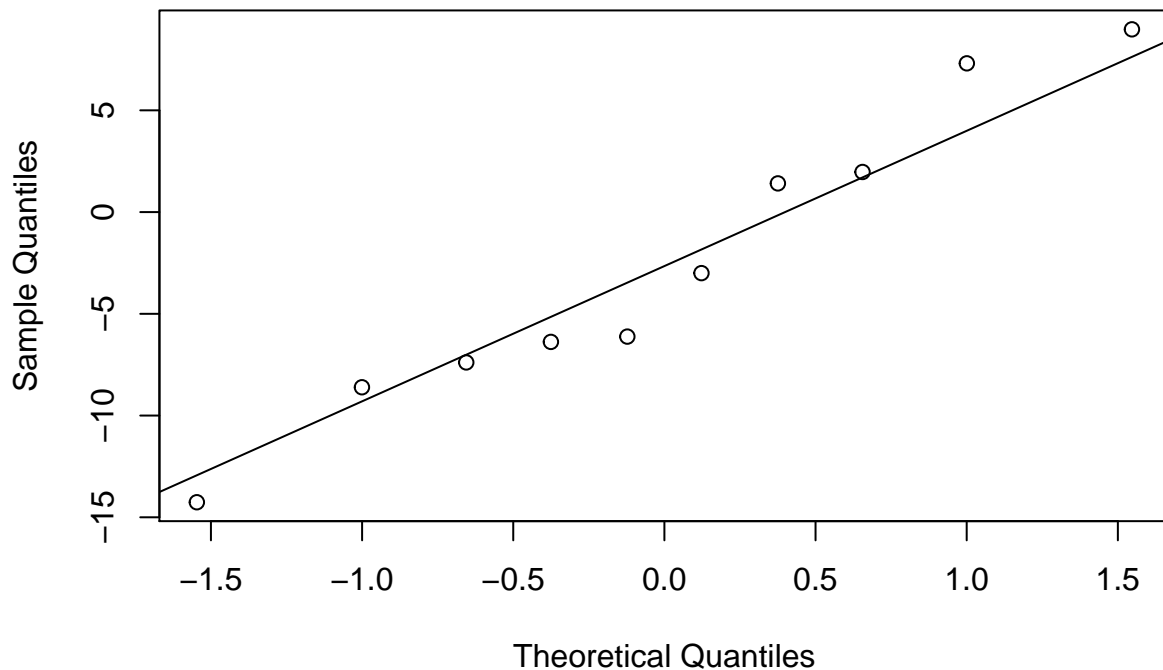
```
hist(delta_but_pro_wide_LOODAT_dbup$delta_propionate)
```

Histogram of delta_but_pro_wide_LOODAT_dbup$delta_propionate

delta_but_pro_wide_LOODAT_dbup$delta_propionate

```r
qqnorm(delta_but_pro_wide_LOODAT_dbup$delta_propionate)
qqline(delta_but_pro_wide_LOODAT_dbup$delta_propionate) #p=0.729. Normality supported
```

Normal Q–Q Plot


```
#Equality of Variances
var.test(delta_but_pro_wide_LOODAT_dbup$delta_butyrate, delta_but_pro_wide_LOODAT_dbup$delta_propionate
```

```
##
##  F test to compare two variances
##
## data:  delta_but_pro_wide_LOODAT_dbup$delta_butyrate and delta_but_pro_wide_LOODAT_dbup$delta_propion
## F = 0.13425, num df = 9, denom df = 9, p-value = 0.006273
## alternative hypothesis: true ratio of variances is not equal to 1
## 90 percent confidence interval:
##   0.04223278 0.42677754
## sample estimates:
## ratio of variances
##           0.1342535
```

```
var.test(delta_but_pro_wide_LOODAT_dbup$delta_butyrate, delta_but_pro_wide_LOODAT_dbup$delta_propionate
```

```
##
##  F test to compare two variances
##
## data:  delta_but_pro_wide_LOODAT_dbup$delta_butyrate and delta_but_pro_wide_LOODAT_dbup$delta_propion
## F = 0.13425, num df = 9, denom df = 9, p-value = 0.9969
## alternative hypothesis: true ratio of variances is greater than 1
## 90 percent confidence interval:
##   0.05501425        Inf
```

```
## sample estimates:
## ratio of variances
##           0.1342535
```

```r
var.test(delta_but_pro_wide_LOODAT_dbup$delta_butyrate, delta_but_pro_wide_LOODAT_dbup$delta_propionate
```

```
##
##  F test to compare two variances
##
## data:  delta_but_pro_wide_LOODAT_dbup$delta_butyrate and delta_but_pro_wide_LOODAT_dbup$delta_propion
## F = 0.13425, num df = 9, denom df = 9, p-value = 0.003136
## alternative hypothesis: true ratio of variances is less than 1
## 90 percent confidence interval:
##  0.0000000 0.3276243
## sample estimates:
## ratio of variances
##           0.1342535
```

```r
#----------------------------------------------------------------------------------
#Assumptions selecting for propionate Conditions:
delta_but_pro_wide_LOODAT_pdown <- delta_but_pro_wide_LOODAT %>%
  filter(delta_propionate < 0) %>%
  summarise(sample_size = n()) #n=9
delta_but_pro_wide_LOODAT_pdown
```

```
## # A tibble: 1 x 1
##   sample_size
##         <int>
## 1           9
```

```r
delta_but_pro_wide_LOODAT_pdown <- delta_but_pro_wide_LOODAT %>%
  filter(delta_propionate < 0)

#Normality
#Butyrate:
shapiro.test(delta_but_pro_wide_LOODAT_pdown$delta_butyrate)
```
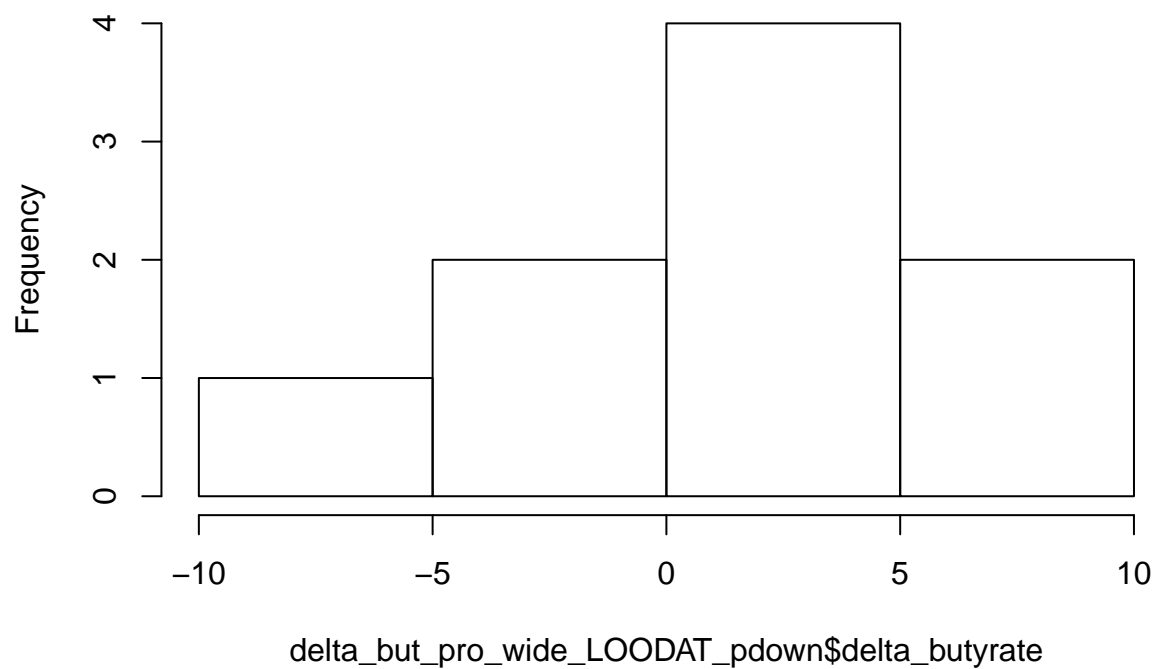
```
##
##  Shapiro-Wilk normality test
##
## data:  delta_but_pro_wide_LOODAT_pdown$delta_butyrate
## W = 0.96189, p-value = 0.8178
```
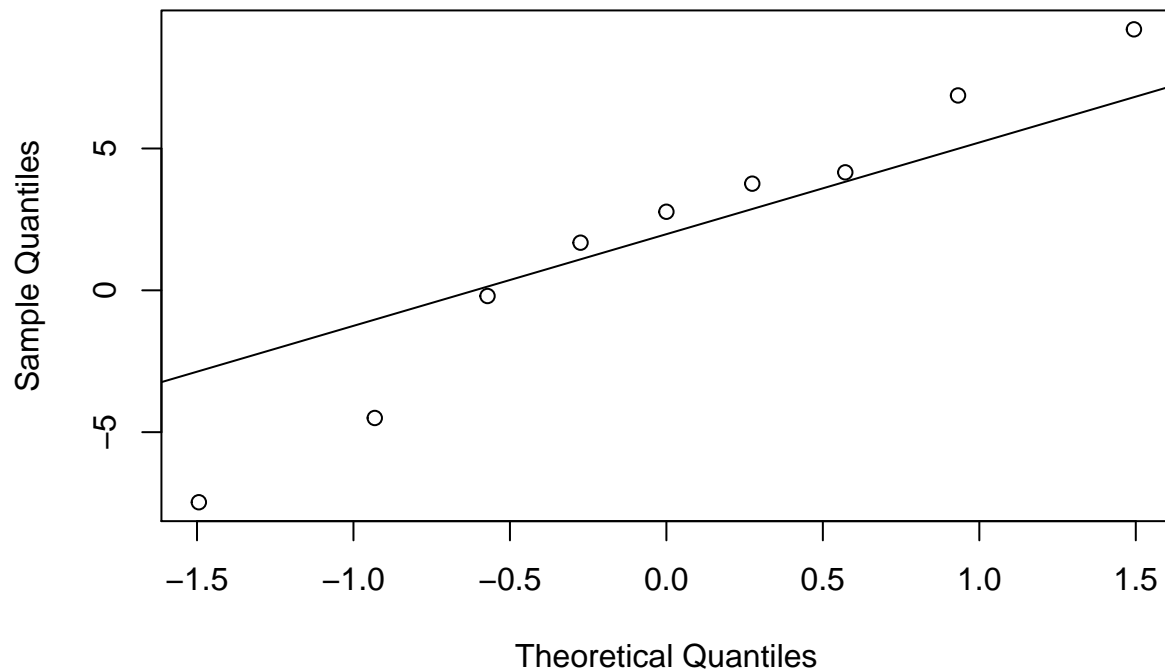
```r
hist(delta_but_pro_wide_LOODAT_pdown$delta_butyrate)
```

Histogram of delta_but_pro_wide_LOODAT_pdown$delta_butyrate

delta_but_pro_wide_LOODAT_pdown$delta_butyrate

```
qqnorm(delta_but_pro_wide_LOODAT_pdown$delta_butyrate)
qqline(delta_but_pro_wide_LOODAT_pdown$delta_butyrate) #p=0.8178. Normality supported.
```
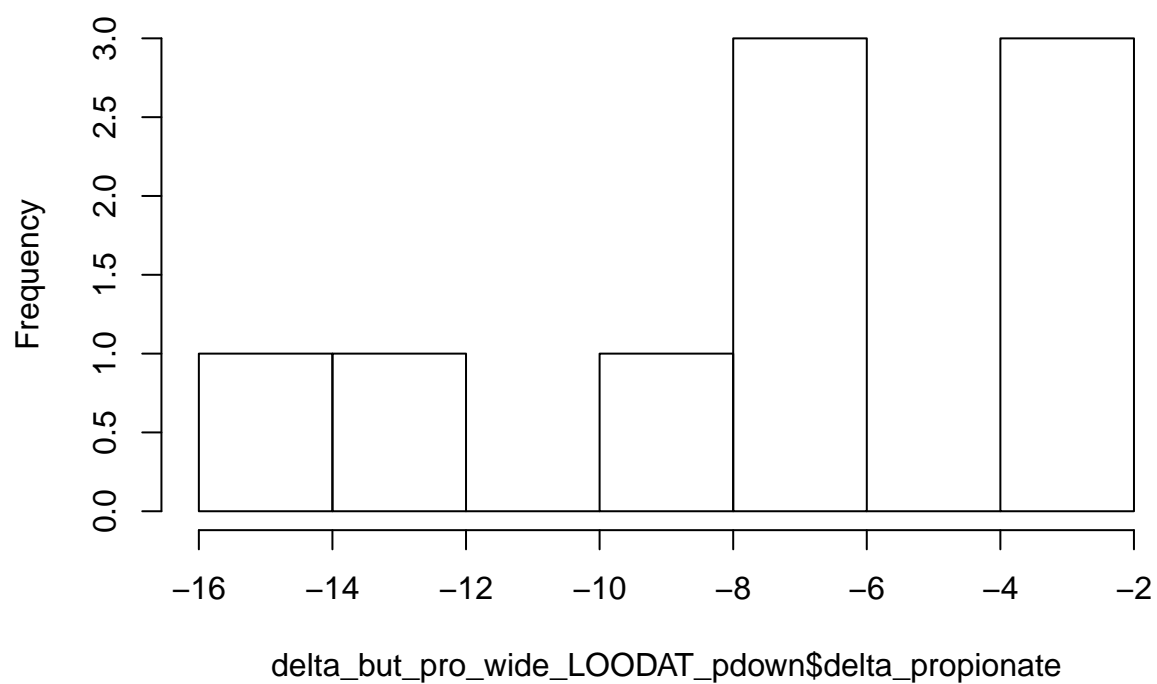
Normal Q–Q Plot


```
#Propionate:
shapiro.test(delta_but_pro_wide_LOODAT_pdown$delta_propionate)
```
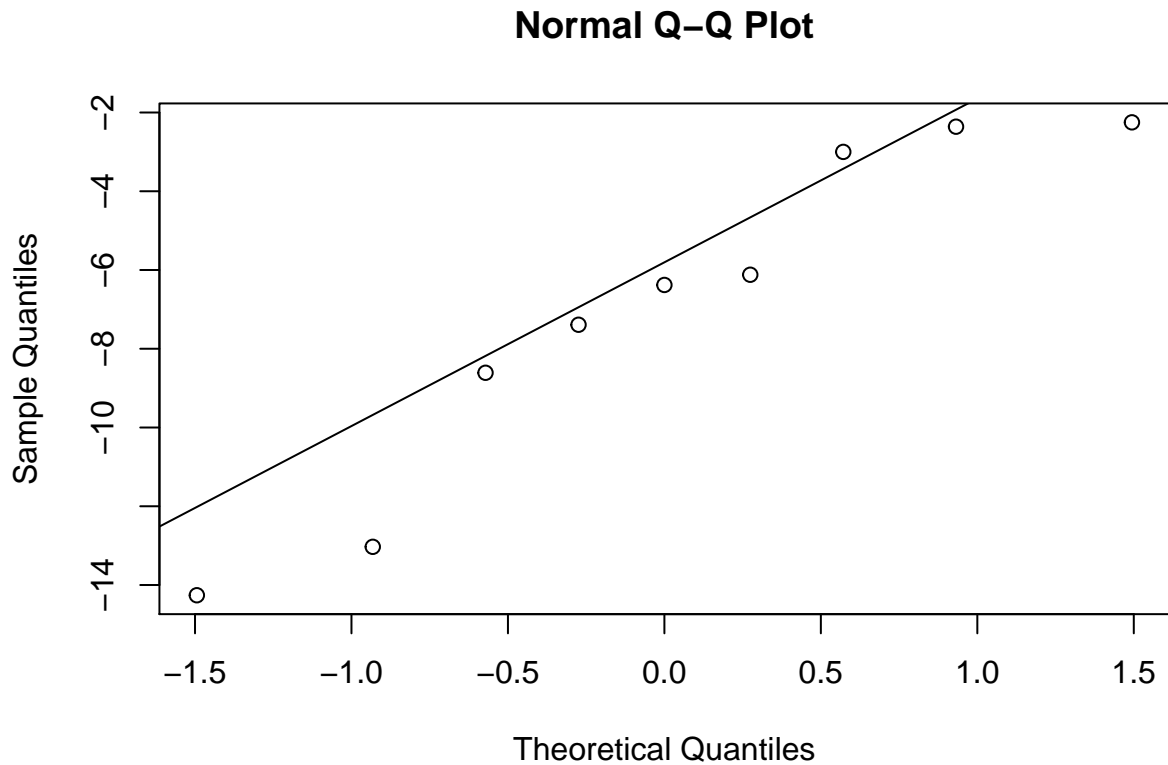
```
##
##  Shapiro-Wilk normality test
##
## data:  delta_but_pro_wide_LOODAT_pdown$delta_propionate
## W = 0.90401, p-value = 0.2762
```

```
hist(delta_but_pro_wide_LOODAT_pdown$delta_propionate)
```

Histogram of delta_but_pro_wide_LOODAT_pdown$delta_propionat

delta_but_pro_wide_LOODAT_pdown$delta_propionate

```
qqnorm(delta_but_pro_wide_LOODAT_pdown$delta_propionate)
qqline(delta_but_pro_wide_LOODAT_pdown$delta_propionate) #p=0.2762. Normality supported
```

Normal Q–Q Plot


```r
#Equality of Variances
var.test(delta_but_pro_wide_LOODAT_pdown$delta_butyrate, delta_but_pro_wide_LOODAT_pdown$delta_propiona
```

```
##
##  F test to compare two variances
##
## data:  delta_but_pro_wide_LOODAT_pdown$delta_butyrate and delta_but_pro_wide_LOODAT_pdown$delta_prop:
## F = 1.4404, num df = 8, denom df = 8, p-value = 0.6178
## alternative hypothesis: true ratio of variances is not equal to 1
## 90 percent confidence interval:
##  0.4189646 4.9523882
## sample estimates:
## ratio of variances
##            1.440443
```

```r
var.test(delta_but_pro_wide_LOODAT_pdown$delta_butyrate, delta_but_pro_wide_LOODAT_pdown$delta_propiona
```

```
##
##  F test to compare two variances
##
## data:  delta_but_pro_wide_LOODAT_pdown$delta_butyrate and delta_but_pro_wide_LOODAT_pdown$delta_prop:
## F = 1.4404, num df = 8, denom df = 8, p-value = 0.3089
## alternative hypothesis: true ratio of variances is greater than 1
## 90 percent confidence interval:
##  0.5562953       Inf
```

```
## sample estimates:
## ratio of variances
##          1.440443
```
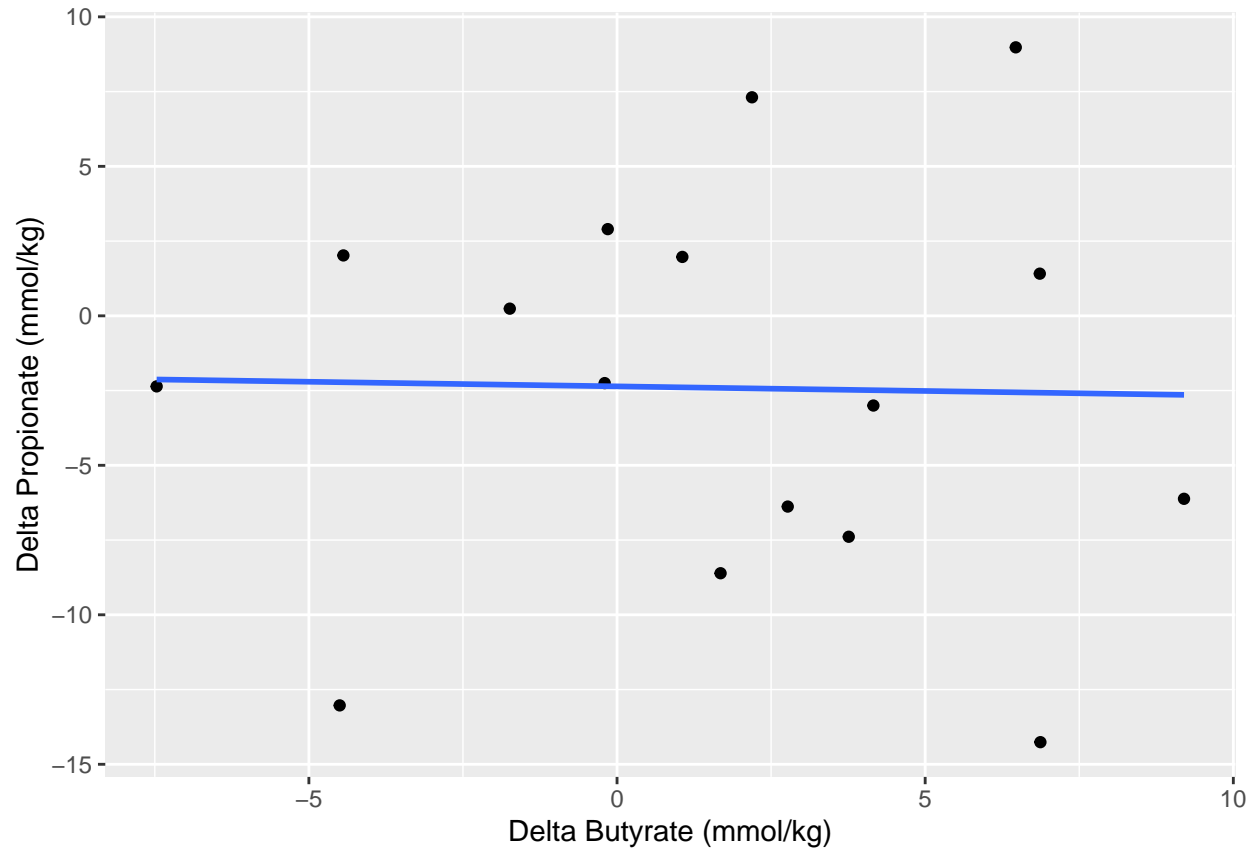
```r
var.test(delta_but_pro_wide_LOODAT_pdown$delta_butyrate, delta_but_pro_wide_LOODAT_pdown$delta_propiona
```

```
##
##  F test to compare two variances
##
## data:  delta_but_pro_wide_LOODAT_pdown$delta_butyrate and delta_but_pro_wide_LOODAT_pdown$delta_prop
## F = 1.4404, num df = 8, denom df = 8, p-value = 0.6911
## alternative hypothesis: true ratio of variances is less than 1
## 90 percent confidence interval:
##   0.000000 3.729809
## sample estimates:
## ratio of variances
##          1.440443
```
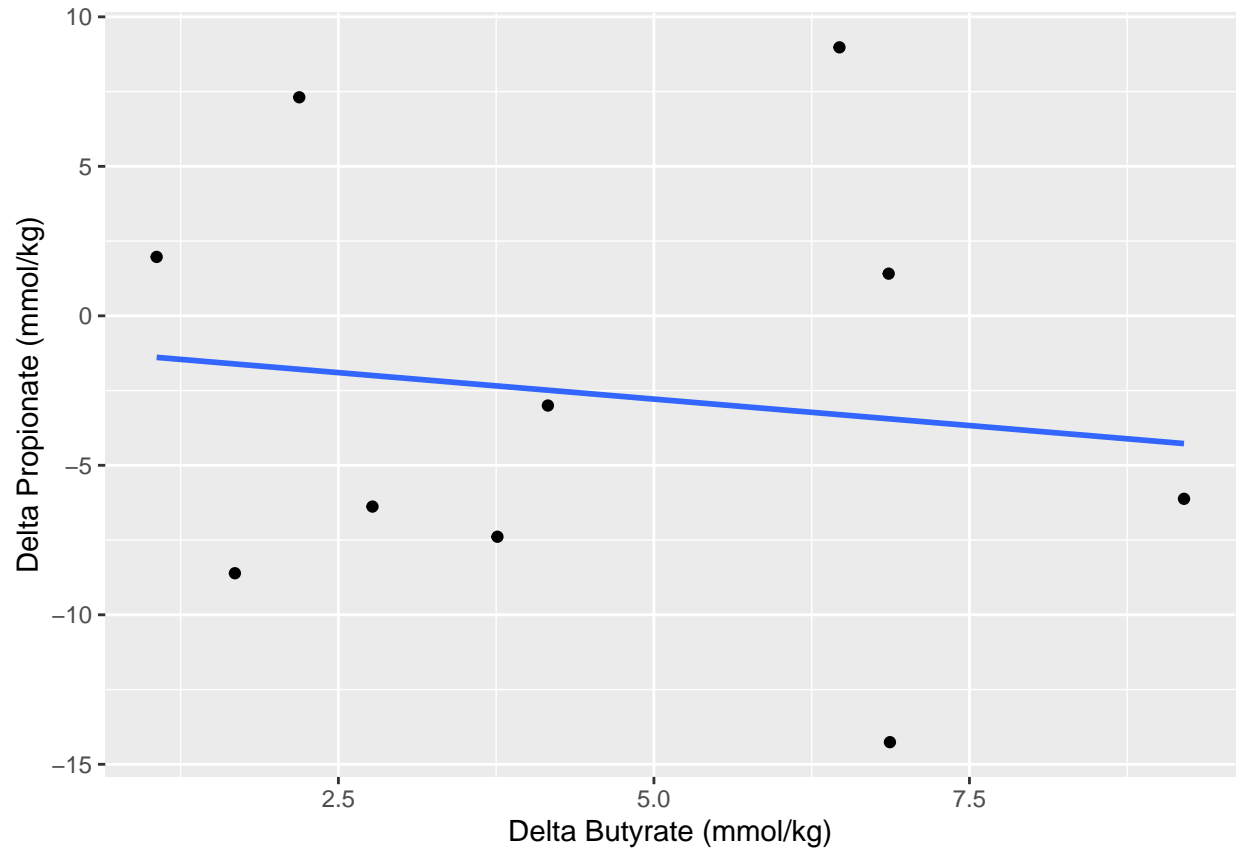
####Plot

```r
LOODAT_correlation_plot <- delta_but_pro_wide_LOODAT %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  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("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(LOODAT_correlation_plot)
```

```
#-------------------------------------------------------------------------------
#Plot Assuming butyrate Conditions Met:
LOODAT_correlation_plot_dbup <- delta_but_pro_wide_LOODAT_dbup %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  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("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(LOODAT_correlation_plot_dbup)
```

```
#-------------------------------------------------------------------------------
#Plot Assuming propionate Conditions Met:
LOODAT_correlation_plot_pdown <- delta_but_pro_wide_LOODAT_pdown %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  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("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(LOODAT_correlation_plot_pdown)
```
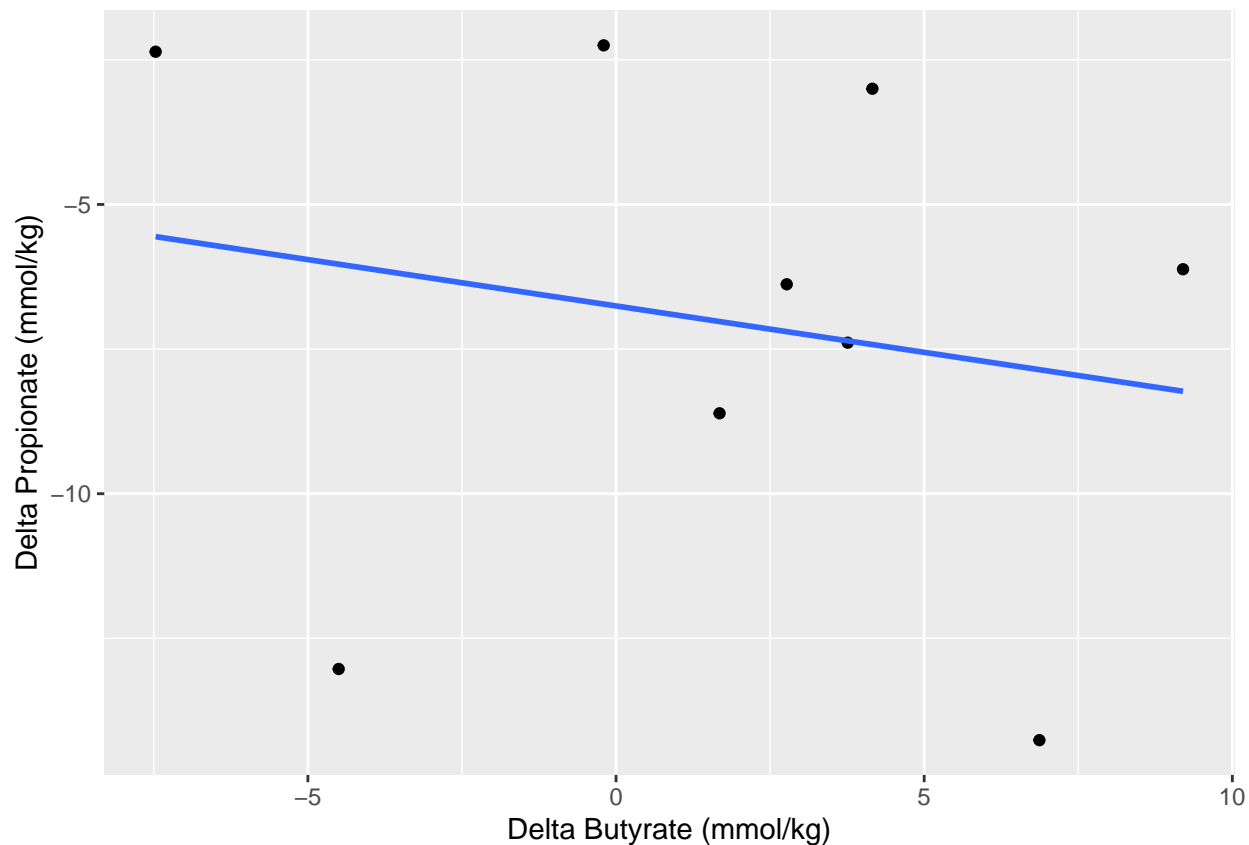
Test

```
LOODAT_test <- delta_but_pro_wide_LOODAT %>%
  lm(delta_propionate ~ delta_butyrate, data = .)
summary(LOODAT_test) #view results
```

```
##
## Call:
## lm(formula = delta_propionate ~ delta_butyrate, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.6886 -4.1799 -0.0638  4.2724  11.5391
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.35949    1.82785  -1.291    0.218
## delta_butyrate -0.03085    0.38141  -0.081    0.937
##
## Residual standard error: 6.86 on 14 degrees of freedom
## Multiple R-squared:  0.0004672,  Adjusted R-squared:  -0.07093
## F-statistic: 0.006543 on 1 and 14 DF,  p-value: 0.9367
```

```
#-------------------------------------------------------------------------------
#Test Assuming Butyrate Condition Met:
LOODAT_test_dbup <- delta_but_pro_wide_LOODAT_dbup %>%
```

```
    lm(delta_propionate ~ delta_butyrate, data = .)
summary(LOODAT_test_dbup) #view results
```

```
##
## Call:
## lm(formula = delta_propionate ~ delta_butyrate, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.812  -4.879  -1.180   4.481  12.286
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.0149     4.9587  -0.205    0.843
## delta_butyrate -0.3541     0.9568  -0.370    0.721
##
## Residual standard error: 7.769 on 8 degrees of freedom
## Multiple R-squared:  0.01683,   Adjusted R-squared:  -0.1061
## F-statistic: 0.137 on 1 and 8 DF,  p-value: 0.7209
```

```
#-------------------------------------------------------------------------------
#Test Assuming Propionate Condition Met:
LOODAT_test_pdown <- delta_but_pro_wide_LOODAT_pdown %>%
  lm(delta_propionate ~ delta_butyrate, data = .)
summary(LOODAT_test_pdown) #view results
```

```
##
## Call:
## lm(formula = delta_propionate ~ delta_butyrate, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9969 -1.5860  0.8187  3.1969  4.4725
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -6.7546     1.6283  -4.148   0.0043 **
## delta_butyrate -0.1603     0.3090  -0.519   0.6199
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.588 on 7 degrees of freedom
## Multiple R-squared:  0.03703,   Adjusted R-squared:  -0.1005
## F-statistic: 0.2692 on 1 and 7 DF,  p-value: 0.6199
```

##BRMPS+HiMaize

```
but_wide_HMB <- all_data_wkly %>%
  rename_all(tolower) %>%
  select(participant_id, semester, supplement_consumed, frequency, quantity_compliant, study_week, butyr
  filter(study_week == "week1" | study_week == "week3", quantity_compliant == "yes", frequency == "1xda
  spread(study_week, butyrate_mean) %>%
```

```
  mutate(delta_butyrate = week3 - week1) #calculate difference in butyrate

pro_wide_HMB <- all_data_wkly %>%
  rename_all(tolower) %>%
  select(participant_id, semester, supplement_consumed, frequency, quantity_compliant, study_week, prop
  filter(study_week == "week1" | study_week == "week3", quantity_compliant == "yes", frequency == "1xda
  spread(study_week, propionate_mean) %>%
  mutate(delta_propionate = week3 - week1) #calculate difference in propionate

delta_but_pro_wide_HMB <- inner_join(x = but_wide_HMB, y = pro_wide_HMB,
                   by = c("participant_id",
                          "semester", "supplement_consumed","frequency", "quantity_compliant")) %>%
  select(-starts_with("week"))
```

Tidy Data ### Analysis #### Assumptions

```
nrow(delta_but_pro_wide_HMB) # number of data points
```

```
## [1] 23
```

```
shapiro.test(delta_but_pro_wide_HMB$delta_butyrate) #call column of df with values = vector
```

```
##
##  Shapiro-Wilk normality test
##
## data:  delta_but_pro_wide_HMB$delta_butyrate
## W = 0.94331, p-value = 0.2767
```

```
shapiro.test(delta_but_pro_wide_HMB$delta_propionate) #call column of df with values = vector
```

```
##
##  Shapiro-Wilk normality test
##
## data:  delta_but_pro_wide_HMB$delta_propionate
## W = 0.88154, p-value = 0.0156
```

```
ggplot(delta_but_pro_wide_HMB, aes(x=delta_butyrate)) +
  geom_histogram() #histograms show the number of data points (count) at each value
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
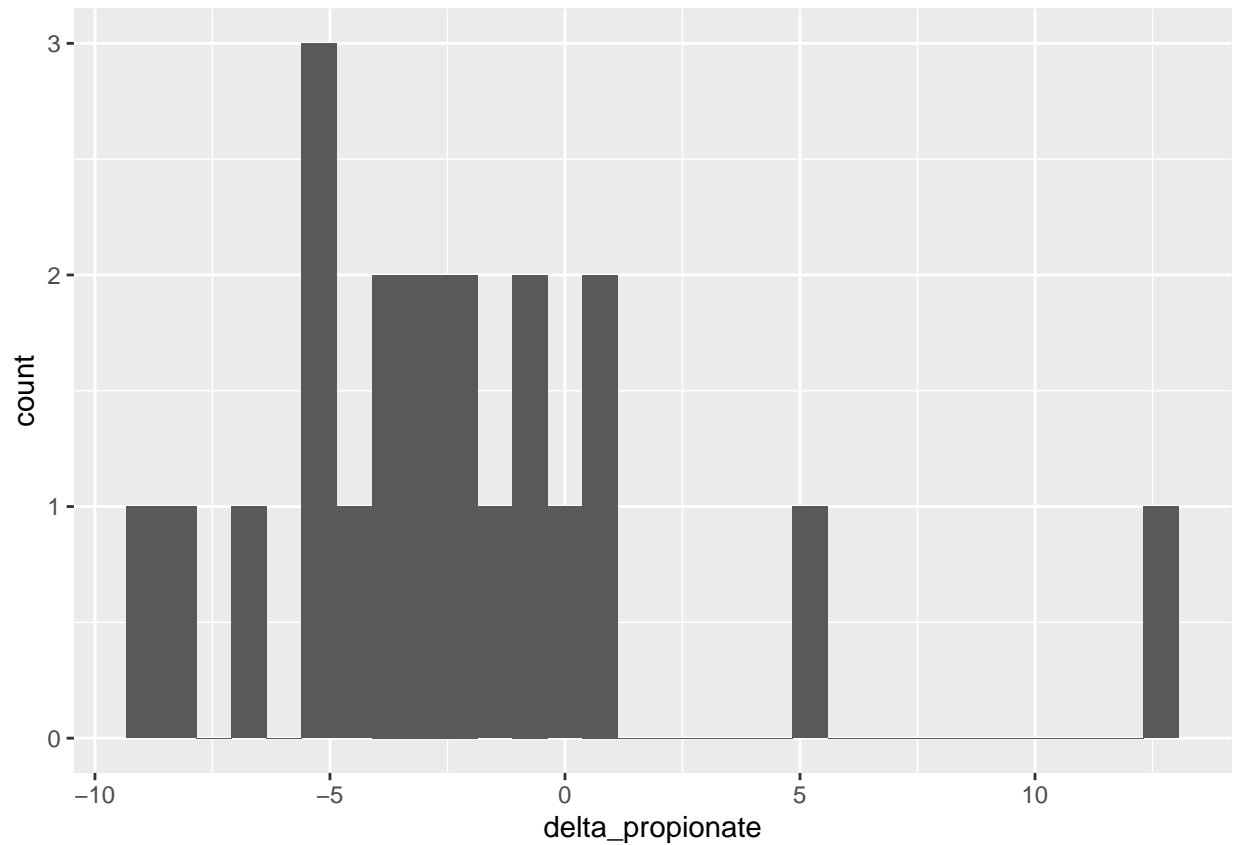
```
## Warning: Removed 3 rows containing non-finite values (stat_bin).
```
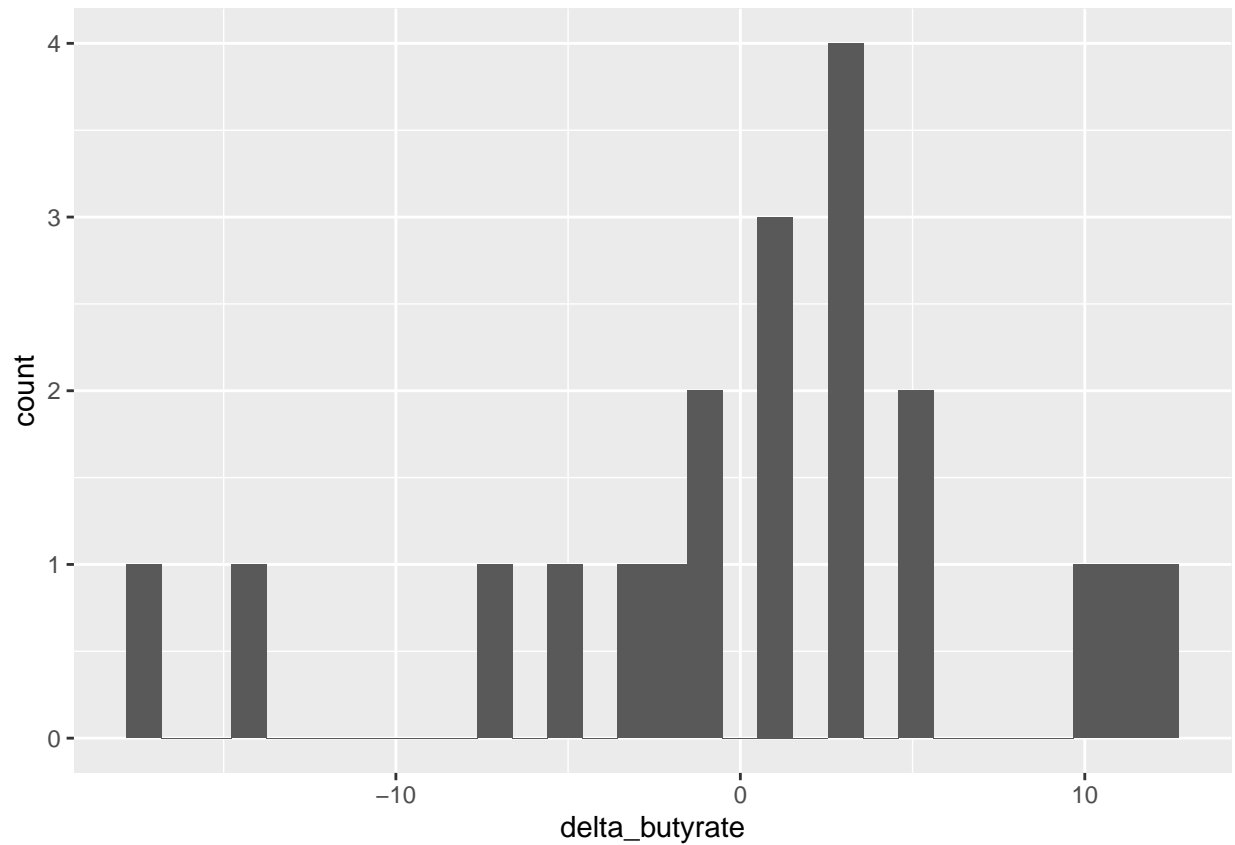
```
ggplot(delta_but_pro_wide_HMB, aes(x=delta_propionate)) +
  geom_histogram() #histograms show the number of data points (count) at each value
```

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

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

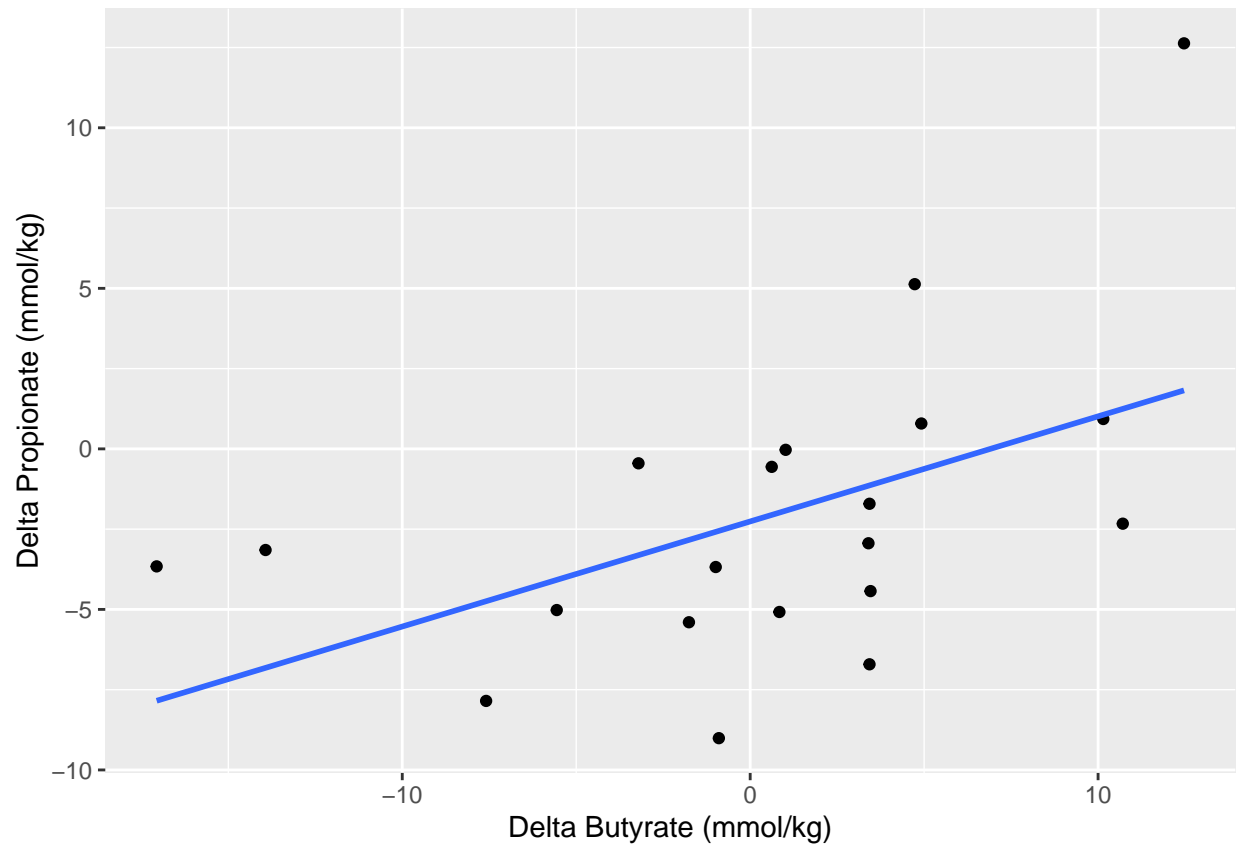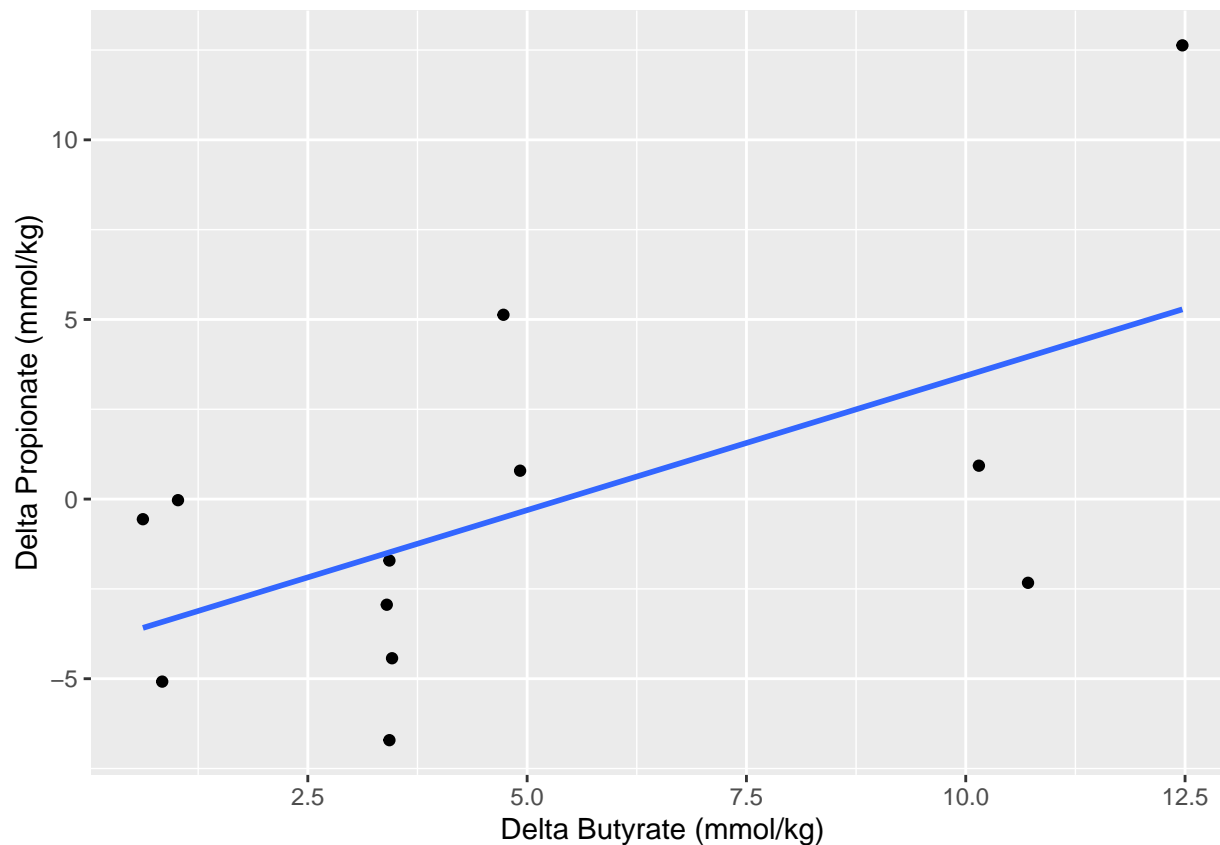#### Plot

```r
linear_correlation_HMB_plot<-delta_but_pro_wide_HMB %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  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("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(linear_correlation_HMB_plot)
```

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

```
## Warning: Removed 3 rows containing missing values (geom_point).
```

```
lm_HMB_highBut_plot<-delta_but_pro_wide_HMB %>%
  filter(delta_butyrate > "0") %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  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("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(lm_HMB_highBut_plot)
```

```
lm_HMB_LowPro_plot<-delta_but_pro_wide_HMB %>%
  filter(delta_propionate < "0") %>%
  ggplot(aes(x = delta_butyrate,
             y = delta_propionate)) +
  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("Delta Butyrate (mmol/kg)") +
  ylab("Delta Propionate (mmol/kg)")
print(lm_HMB_LowPro_plot)
```

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

Warning: Removed 1 rows containing missing values (geom_point).
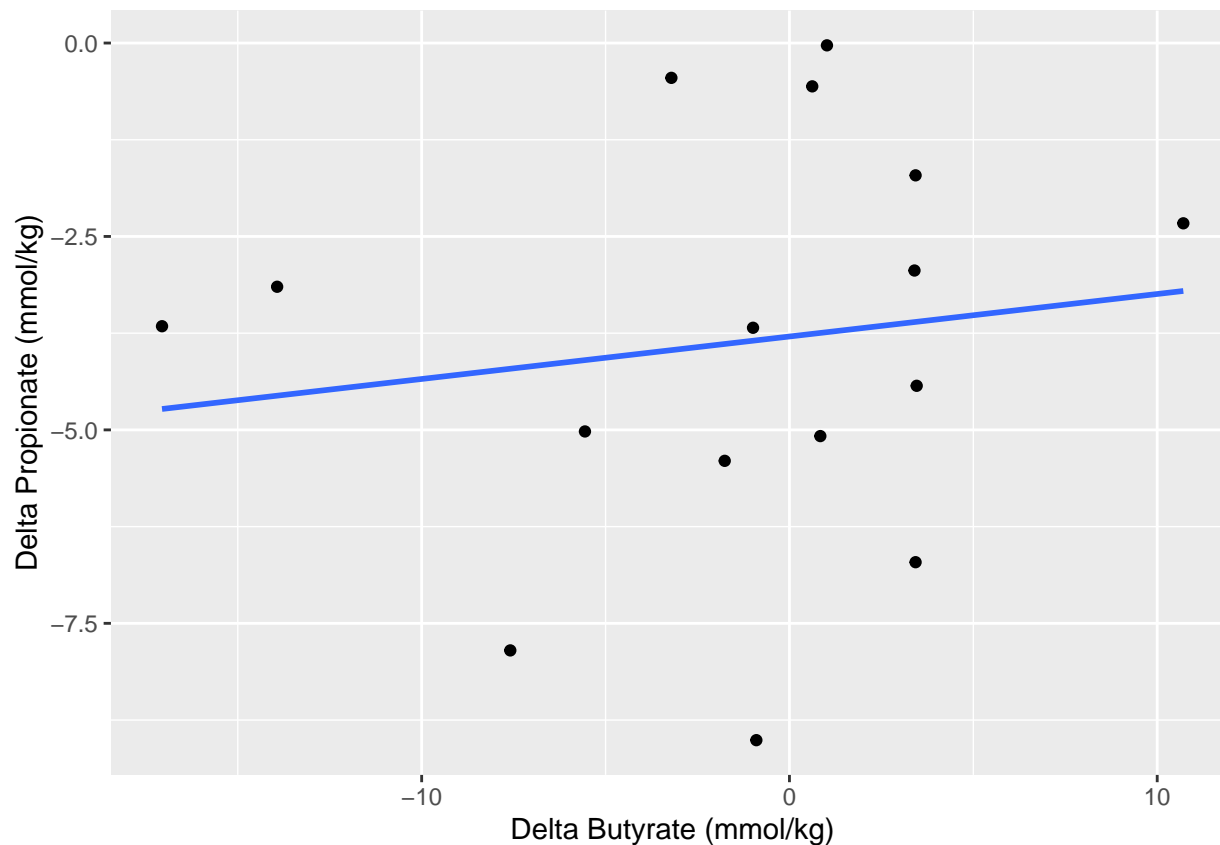
Test

```
HMB_lm <- delta_but_pro_wide_HMB %>%
  lm(delta_butyrate ~ delta_propionate, data = .) #test relationship
summary(HMB_lm) #view results
```

```
##
## Call:
## lm(formula = delta_butyrate ~ delta_propionate, data = .)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -16.2628  -1.9142   0.4301   3.7494  10.4611
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.0816     1.6196   1.285   0.2150
## delta_propionate   0.7865     0.3148   2.498   0.0224 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.595 on 18 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.2575, Adjusted R-squared:  0.2162
## F-statistic: 6.242 on 1 and 18 DF,  p-value: 0.02238
```

```r
# Small .022 p-value indicates significance, and the other R2 is small-ish at .2 representing that the

HMB_lm_HiBut <- delta_but_pro_wide_HMB %>%
  filter(delta_butyrate > "0") %>%
  lm(delta_butyrate ~ delta_propionate, data = .) #test relationship
summary(HMB_lm_HiBut) #view results
```

```
##
## Call:
## lm(formula = delta_butyrate ~ delta_propionate, data = .)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2210 -2.1401 -0.4485  1.4428  6.6682
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.0938     0.9913   5.139 0.000439 ***
## delta_propionate   0.4515     0.1999   2.259 0.047472 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.425 on 10 degrees of freedom
## Multiple R-squared:  0.3378, Adjusted R-squared:  0.2716
## F-statistic: 5.102 on 1 and 10 DF,  p-value: 0.04747
```

```r
# Small .047 p-value indicates significance, and the  R2 is small-ish at .33 representing that the corr

HMB_lm_LoPro <- delta_but_pro_wide_HMB %>%
  filter(delta_propionate < "0") %>%
  lm(delta_butyrate ~ delta_propionate, data = .) #test relationship
summary(HMB_lm_LoPro) #view results
```

```
##
## Call:
## lm(formula = delta_butyrate ~ delta_propionate, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.6378  -3.1755   0.9399   4.2090  11.6175
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.005734   3.245037  -0.002    0.999
## delta_propionate  0.387006   0.701894   0.551    0.590
##
## Residual standard error: 7.077 on 14 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.02125,   Adjusted R-squared:  -0.04866
## F-statistic: 0.304 on 1 and 14 DF,  p-value: 0.5901
```

Somewhat small .59 p-value indicates possible significance, and the tiny R2 of -.049 represents that