Contents

Incubation

Masters experiment, incubation of soils amended with various amendments

Load libraries

```
library(phyloseq)
library(tidyverse)
library(vegan)
library(ggpubr)
```

Read in the data

use `readRDS` to load phyloseq object

```
inc.raw <- readRDS("Data/incubation_raw.RDS")
inc.raw
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 43726 taxa and 350 samples ]
## sample_data() Sample Data:       [ 350 samples by 13 sample variables ]
## tax_table()   Taxonomy Table:    [ 43726 taxa by 6 taxonomic ranks ]
```

Functions

```
# Put phyloseq object into a df with .02% phylum (glomed at phylum level)
RelativeAbundanceDf <- function(physeq) {
    physeq %>% tax_glom(taxrank = "Phylum") %>% transform_sample_counts(function(x) {
        x/sum(x)
```

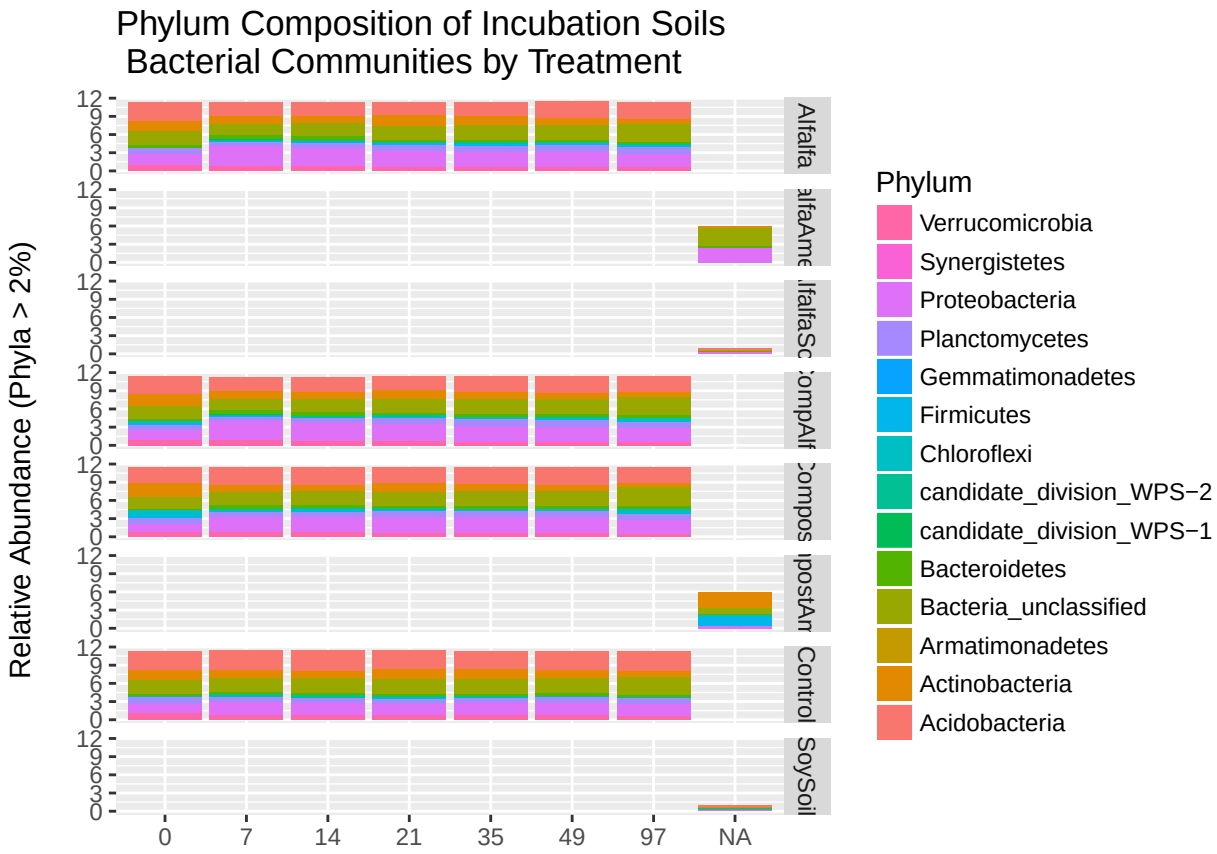
```
    }) %>% psmelt() %>% filter(Abundance > 0.02) %>% arrange(Phylum)
}

# Function to plot relative abundance
PlotRelativeAbundance <- function(df) {
    ggplot(df, aes(x = as.factor(day), y = Abundance, fill = Phylum)) +
    facet_grid(treatment ~ .) +
    geom_bar(stat = "identity") +
    #scale_fill_manual(values = phylum.colors) +
        # Remove x axis title
    theme(axis.title.x = element_blank()) +
    guides(fill = guide_legend(reverse = TRUE, keywidth = 1, keyheight = 1)) +
    ylab("Relative Abundance (Phyla > 2%) \n") +
    ggtitle("Phylum Composition of Incubation Soils \n Bacterial Communities by Treatment")
}

#Scale reads function to be used prior to ordination
ScaleReads <- function(physeq, n) {
  physeq.scale <- transform_sample_counts(physeq, function(x) {
    (n * x/sum(x))
  })
  otu_table(physeq.scale) <- floor(otu_table(physeq.scale))
  physeq.scale <- prune_taxa(taxa_sums(physeq.scale) > 0, physeq.scale)
  return(physeq.scale)
}

# Function to summarise a data frame and give statistics
DataSummary <- function(data, varname, groupnames) {
  require(plyr)
  SummaryFunc <- function(x, col) {
    c(mean = mean(x[[col]], na.rm = TRUE), sd = sd(x[[col]], na.rm = TRUE))
  }
  data.sum <- ddply(data, groupnames, .fun = SummaryFunc, varname)
  data.sum <- rename(data.sum, c(mean = varname))
}
```

Use function from above to create df with .02% of the phylum level of OTUs

```
inc.raw.phylum.2percent <- RelativeAbundanceDf(inc.raw)

# Plot and save image
inc.phylum.abundance <- PlotRelativeAbundance(inc.raw.phylum.2percent)
tiff('Images/inc.phylum.abundance.tiff', units="in", width=5, height=5, res=300)
inc.phylum.abundance
dev.off()
```

```
## pdf
##   2
```

```
inc.phylum.abundance
```



Some data wrangling to make the plots look nicer

```
# First split into two phyloseq objects
# Incubation
inc.treatment <- subset_samples(inc.raw, day %in% c("0", "7", "14", "21", "35", "49", "97"))
# Amends
inc.amend <- subset_samples(inc.raw, treatment %in% c("AlfalfaAmend", "CompostAmend"))

# Now let's pool the reps so that y-axis goes to 1, need to do for each object
inc.merged <- inc.treatment
variable.1 <- as.character(get_variable(inc.merged, "treatment"))
variable.2 <- as.character(get_variable(inc.merged, "day"))
sample_data(inc.merged)$TreatmentAndDay <- mapply(paste0, variable.1, variable.2, collapse = "-")
inc.merged <- merge_samples(inc.merged, "TreatmentAndDay")
sample_data(inc.merged)$treatment <- levels(sample_data(inc.treatment)$treatment)
```

Relative abundance of phyla

```
# Innie to outtie, make df and plot relative abumndace
treatments.abundance <- PlotRelativeAbundance(RelativeAbundanceDf(inc.treatment))
amendments.abundance <- PlotRelativeAbundance(RelativeAbundanceDf(inc.amend))
mergeddf.abundance <- PlotRelativeAbundance(RelativeAbundanceDf(inc.merged))

# Save as images, high quality
tiff('Images/treatments.abundance.tiff', units="in", width=5, height=5, res=300)
treatments.abundance
```

```
dev.off()
```

```
## pdf
##   2
```

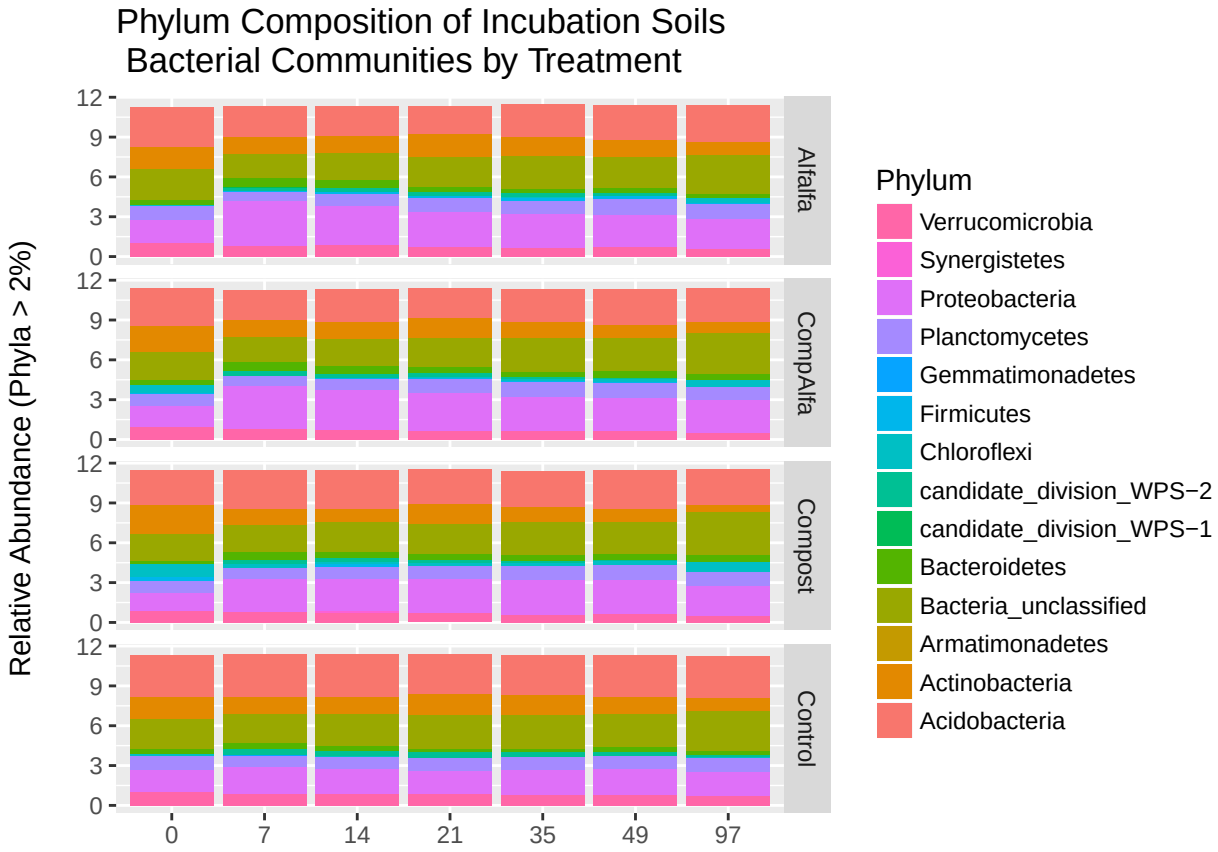
```
tiff('Images/amendments.abundance.tiff', units="in", width=5, height=5, res=300)
amendments.abundance
dev.off()
```

```
## pdf
##   2
```

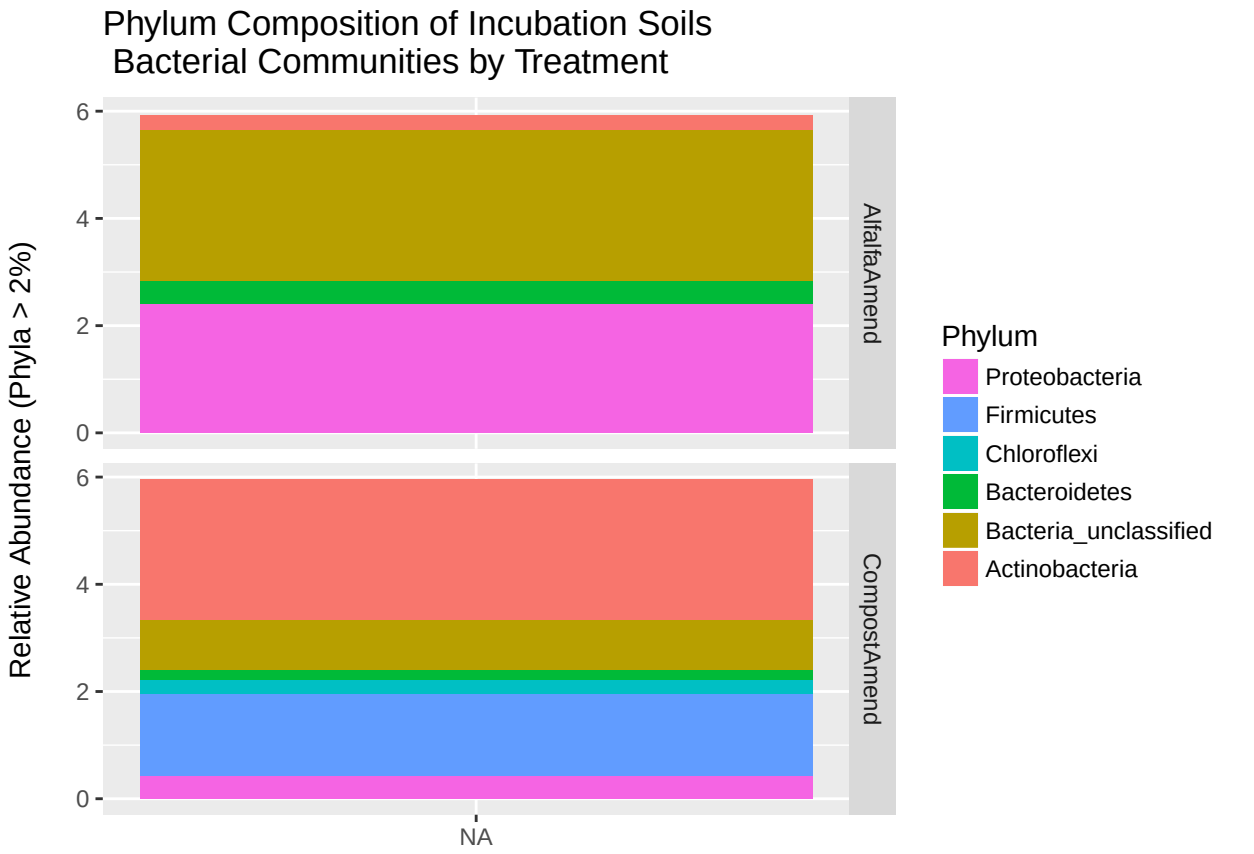
```
tiff('Images/mergeddf.abundance.tiff', units="in", width=5, height=5, res=300)
mergeddf.abundance
dev.off()
```

```
## pdf
##   2
```

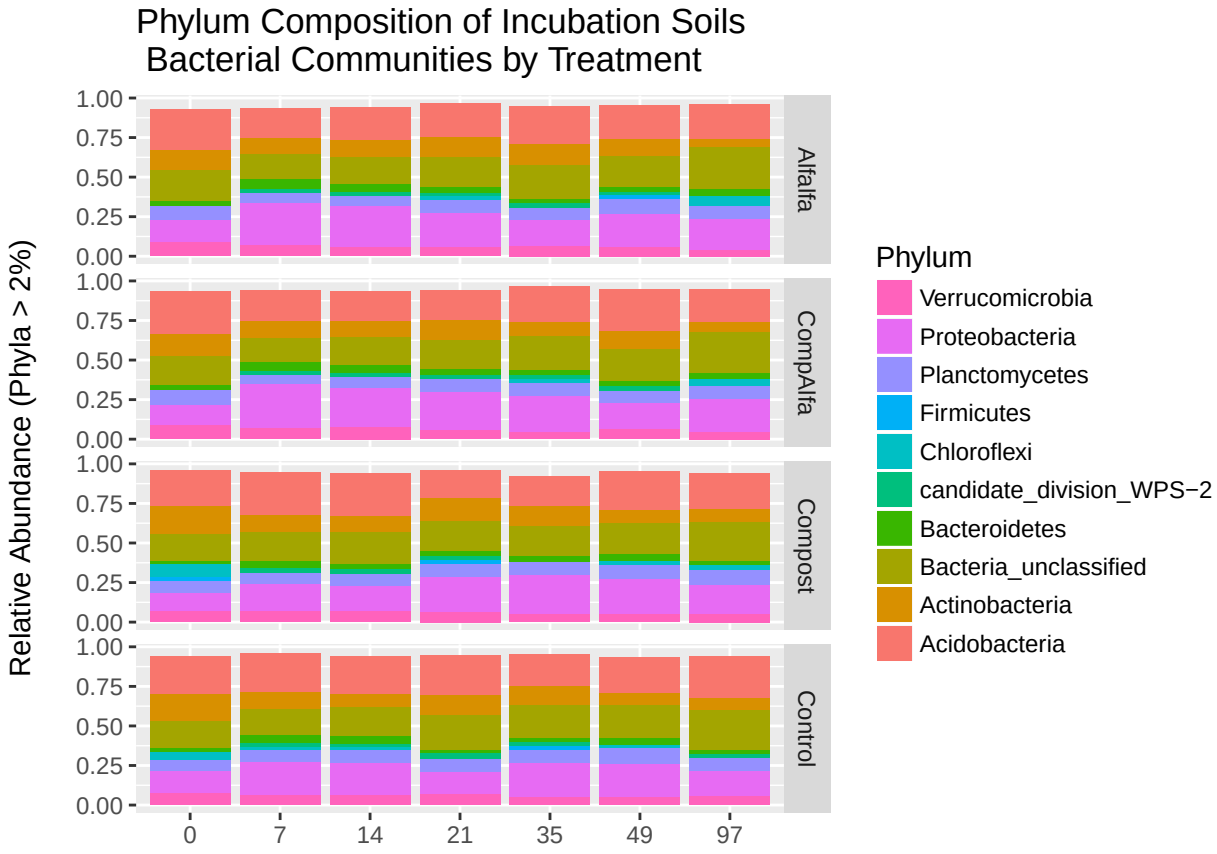
```
treatments.abundance
```




```
amendments.abundance
```




```
mergeddf.abundance
```



PCoA

We already have the function `ScaleReads` to re-sample our data to a specific read number per sample.

```
# Now call the function with the phyloseq object you wish to scale
inc.scale.treatment <- ScaleReads(inc.treatment, n = 6000)
```

Fix day levels in `sample_data`

```
sample_data(inc.scale.treatment)$day <- factor(sample_data(inc.scale.treatment)$day,
  levels = c("0", "7", "14", "21", "35", "49", "97"))
```

Now use the `ordinate` function from phyloseq

```
inc.ordination.treatment <- ordinate(physeq = inc.scale.treatment, method = "PCoA", distance = "bray")
```

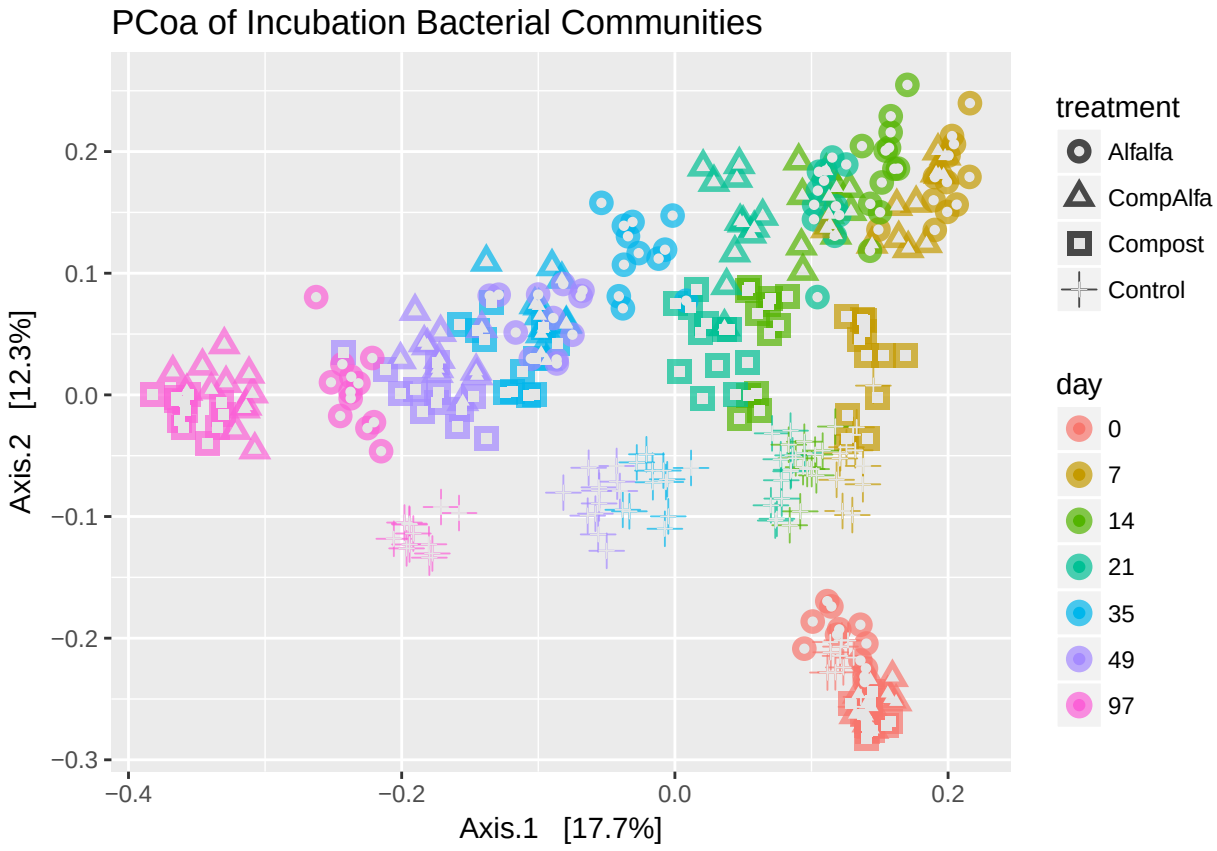
Use `plot_ordination` to create the plot then manipulate it with ggplot2

```
inc.ordination.plot.treatment <- plot_ordination(physeq = inc.scale.treatment, ordination = inc.ordinat
  color = "day", shape = "treatment", title = "PCoa of Incubation Bacterial Communities") +
  #scale_color_manual(values = phylum.colors) +
  geom_point(aes(color = day), alpha = 0.7, size = 4) +
  geom_point(color = "grey90", size = 1.5)

tiff('Images/inc.ordination.plot.treatment.tiff', units="in", width=10, height=10, res=300)
inc.ordination.plot.treatment
dev.off()
```

```
## pdf
```

```
##    2
```

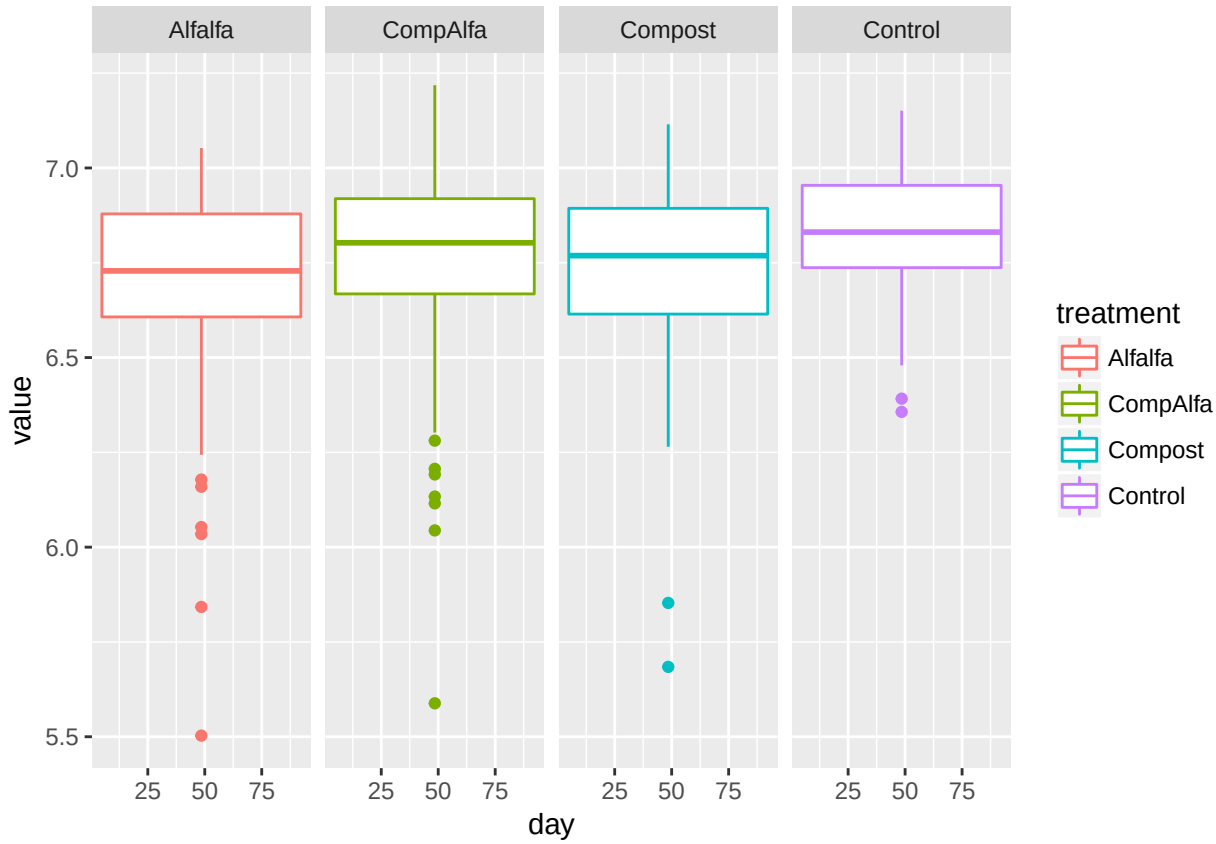
```
inc.ordination.plot.treatment
```



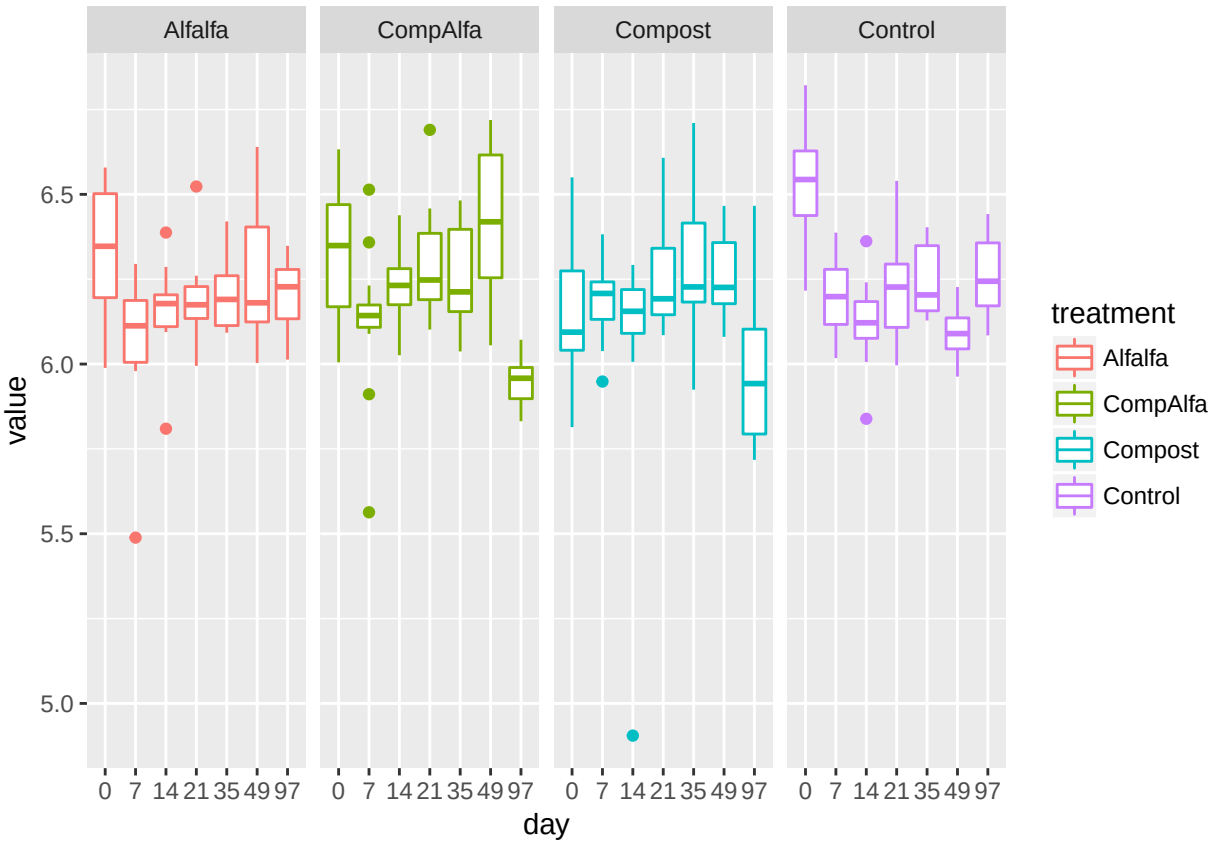
Alpha Diversity

First remove OTUs that sum to 0

```
inc.treatment.pruned <- prune_species(speciesSums(inc.treatment) > 0, inc.treatment)
richness.data <- plot_richness(inc.treatment.pruned, measures = "Shannon")
richness.df <- richness.data$data

scaled.richness.data <- plot_richness(inc.scale.treatment, measures = "Shannon")
scaled.df <- scaled.richness.data$data

richness.shannon <- ggplot(richness.df, aes(x = day, y = value), color = treatment) +
  geom_boxplot(aes(color = treatment), position = "dodge") +
  facet_grid(~treatment)
richness.shannon
```




```
scaled.richness.shannon <- ggplot(scaled.df, aes(x = day, y = value), color = treatment) +
  geom_boxplot(aes(color = treatment), position = "dodge") +
  facet_grid(~treatment)
scaled.richness.shannon
```



Statistics

Interesting results shown so far, we can see that the communities are changing over time and in response to nutrients

```
# Stats
# Adonis
inc.scale.df <- as(sample_data(inc.scale.treatment), "data.frame")
inc.distance <- distance(inc.scale.treatment, "bray")
inc.adonis <- adonis(inc.distance ~ treatment + day, inc.scale.df)
inc.adonis
```

```
##
## Call:
## adonis(formula = inc.distance ~ treatment + day, data = inc.scale.df)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##            Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## treatment   3     4.313 1.43755  17.199 0.09412  0.001 ***
## day         6    14.259 2.37648  28.432 0.31119  0.001 ***
## Residuals 326    27.249 0.08359         0.59469
## Total     335    45.820                 1.00000
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# I'm pretty sure the order you feed the categories
incadonis.reverse <- (adonis(inc.distance ~ day + treatment, inc.scale.df))
```

Soil chemical measurements

Plot the other data, mainly we want to explore how the nutrient concentration and microbial biomass numbers are changing during the course of the incubation.

```
inc.treatment
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 43726 taxa and 336 samples ]
## sample_data() Sample Data:       [ 336 samples by 13 sample variables ]
## tax_table()   Taxonomy Table:    [ 43726 taxa by 6 taxonomic ranks ]
```

```
sample_data(inc.treatment)$day <- as.factor(sample_data(inc.treatment)$day)
inc.data <- as.data.frame(sample_data(inc.treatment))
colnames(inc.data)
```

```
##  [1] "i_id"                      "sample"
##  [3] "treatment"                 "replication"
##  [5] "jar"                       "day"
##  [7] "pH"                        "N_flash"
##  [9] "C_flash"                   "gravimetric_water_content"
## [11] "NH3"                       "NO3"
## [13] "MBC_mg.kg_per_dry_wt_soil"
```

Nitrate "NO3"

```
inc.nitrate.eb <- DataSummary(inc.data, varname = "NO3", groupnames = c("day", "treatment"))
inc.nitrate.eb
```

```
##    day treatment        NO3         sd
## 1    0   Alfalfa        NaN         NA
## 2    0  CompAlfa        NaN         NA
## 3    0   Compost        NaN         NA
## 4    0   Control        NaN         NA
## 5    7   Alfalfa  4.0158333 1.40372827
## 6    7  CompAlfa  0.7890000 0.26608099
## 7    7   Compost  2.3883333 0.41559166
## 8    7   Control  4.3295000 0.45983881
## 9   14   Alfalfa  6.7978333 1.02079528
## 10  14  CompAlfa  1.9499167 0.31778136
## 11  14   Compost  0.1792500 0.09437783
## 12  14   Control  5.7135000 0.37824928
## 13  21   Alfalfa 11.6335622 1.88920063
## 14  21  CompAlfa  4.6493122 0.93634659
## 15  21   Compost  0.0373955 0.03194408
## 16  21   Control  6.6623122 1.23079502
## 17  35   Alfalfa 20.8315490 1.73690734
## 18  35  CompAlfa  9.2763823 1.75712854
## 19  35   Compost  0.1278823 0.09516716
```

```
## 20  35   Control 11.4509657 0.61981764
## 21  49   Alfalfa 28.1442543 4.47671101
## 22  49  CompAlfa 14.4175043 0.70669332
## 23  49   Compost  0.2011710 0.20800801
## 24  49   Control 14.7184210 0.58758991
## 25  97   Alfalfa 38.1015833 0.94668363
## 26  97  CompAlfa 22.2865000 0.93478170
## 27  97   Compost  6.6065000 2.39839062
## 28  97   Control 21.3129167 1.22645813
```

```
plot.inc.nitrate.eb <- ggplot(inc.nitrate.eb, aes(x = day, y = NO3, group = treatment, color = treatmen
  geom_errorbar(aes(ymin = NO3 - sd, ymax = NO3 + sd), width = 1, position = position_dodge(0.05)) +
  geom_line(aes(linetype = treatment)) +
  geom_point(aes(shape = treatment)) +
  labs(title = "Plot of NO3 by day", x = "Day", y = "NO3") +
  theme_classic()
tiff('Images/nitrate.eb.tiff', units="in", width=5, height=5, res=300)
plot.inc.nitrate.eb
dev.off()
```

```
## pdf
##   2
```

```
inc.nitrate <- ggplot(inc.data, aes(x = day, y = NO3, group = treatment)) +
  geom_point(aes(color = treatment))
tiff('Images/nitrate.tiff', units="in", width=5, height=5, res=300)
inc.nitrate
dev.off()
```

```
## pdf
##   2
```

```
plot <- ggplot(inc.data, aes(x = treatment, y = NO3, color = treatment)) +
  facet_grid(~day) +
  geom_boxplot(position = "dodge") +
  theme(axis.title.x=element_blank(), axis.text.x=element_blank(), axis.ticks.x=element_blank())
tiff("Images/boxplot.nitrate.tiff", units = "in", width = 5, height = 5, res = 300)
plot
dev.off()
```

```
## pdf
##   2
```

```
inc.data.7 <- inc.data %>%
  filter(day == 7) %>%
  ggplot(aes(x = treatment, y = NO3, color = treatment)) +
  geom_boxplot() +
  #facet_grid( ~ day) +
  rotate_x_text(angle = 45) +
  geom_hline(yintercept = mean(inc.data$NO3), color = "red") +
  stat_compare_means(method = "anova") +
  stat_compare_means(label = "p.signif", method = "t.test", ref.group = ".all.")
tiff("Images/nitrate.day.7.boxplot.tiff", units = "in", width = 5, height = 5, res = 300)
inc.data.7
dev.off()
```

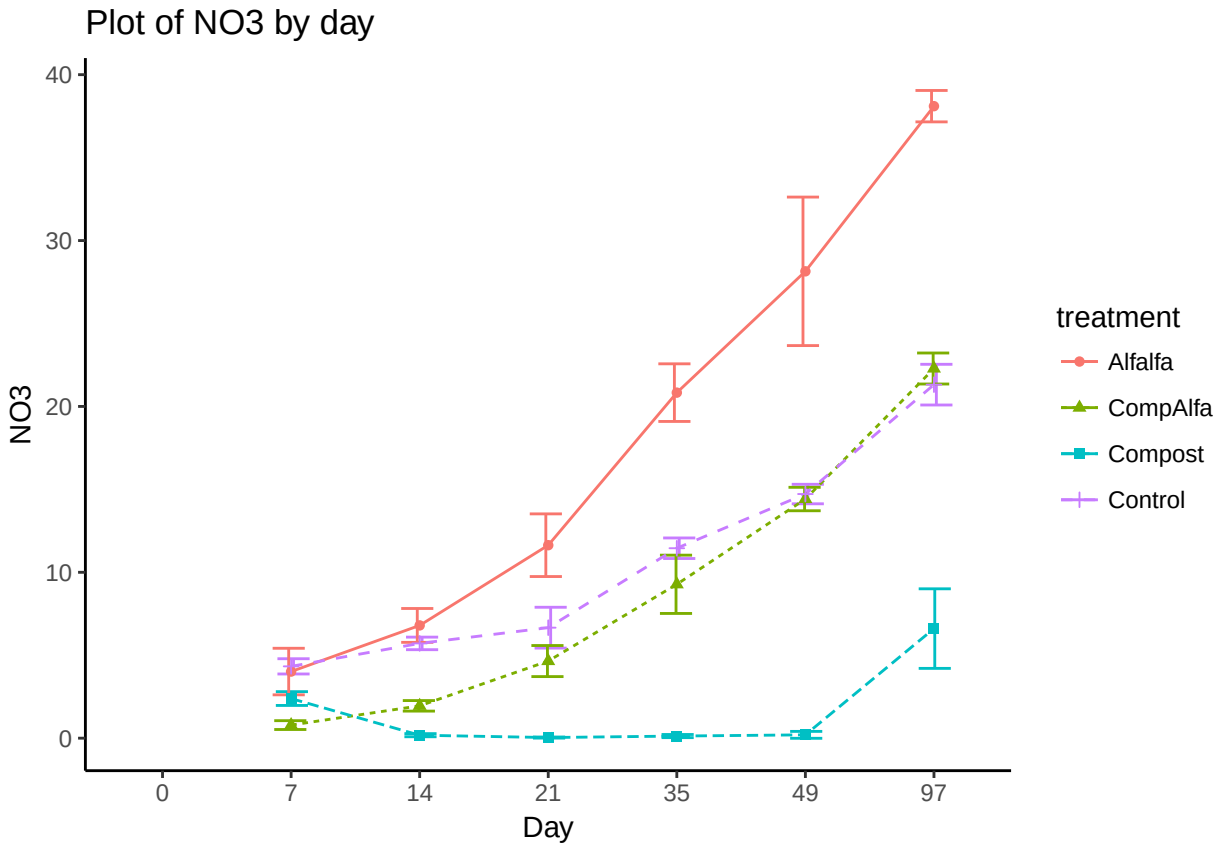
```
## pdf
```

```
##   2
```

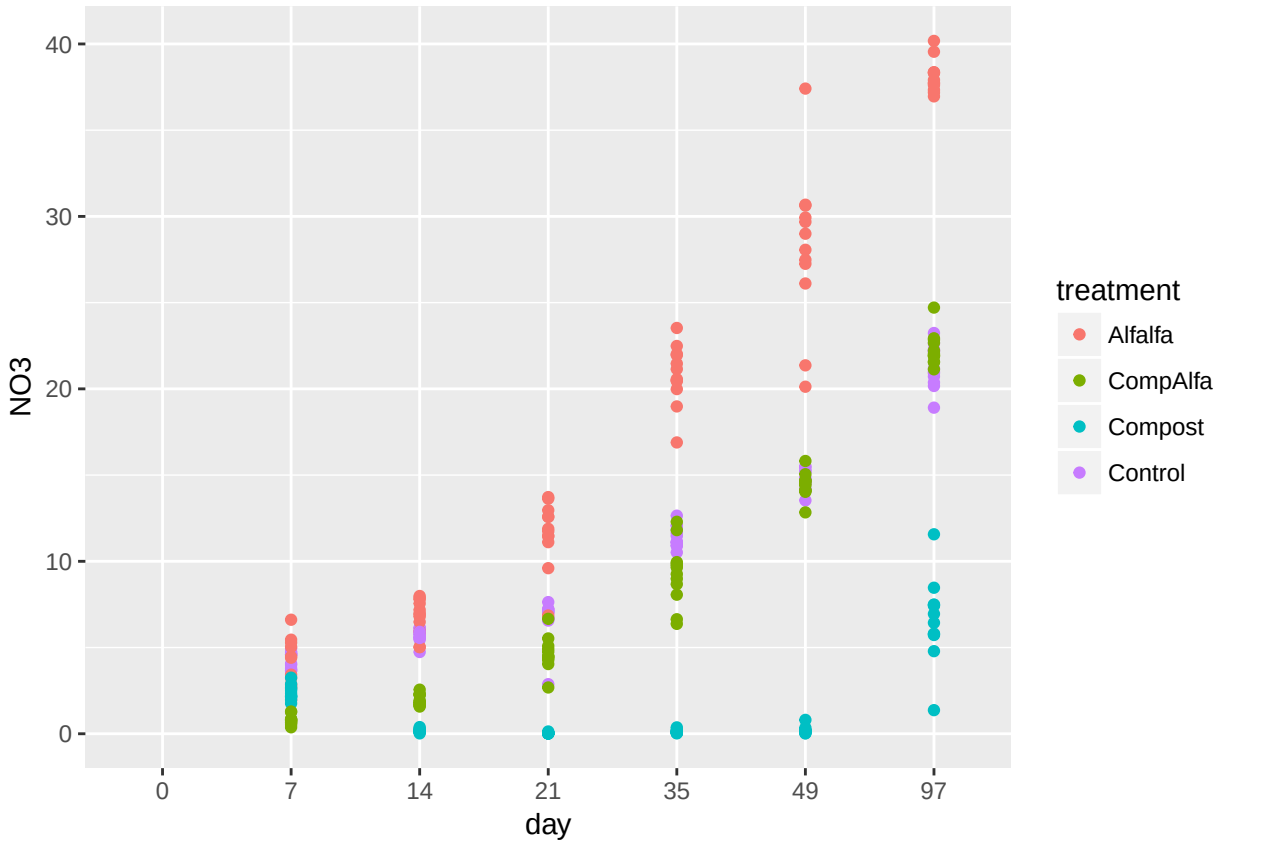
```
test <- inc.data %>%
  filter(day == 7)

stat.day.7 <- ggboxplot(test, x = "treatment", y = "NO3", color = "treatment", legend = "none") +
  rotate_x_text(angle = 45) +
  geom_hline(yintercept = mean(test$NO3)) +
  stat_compare_means(method = "anova") +
  stat_compare_means(label = "p.signif", method = "t.test", ref.group = ".all.")
tiff('Images/stat.day.7.tiff', units="in", width=5, height=5, res=300)
#insert ggplot code
stat.day.7
dev.off()
```

```
## pdf
##   2
```

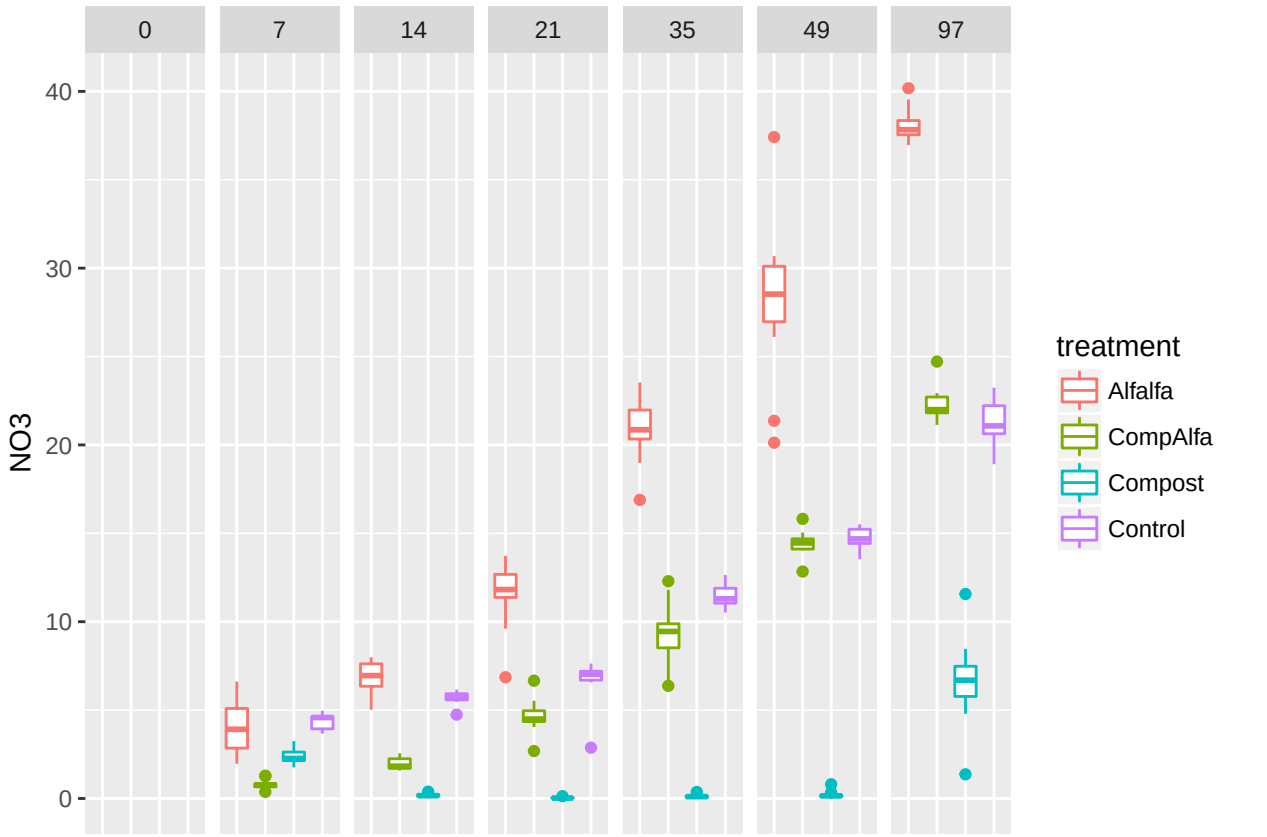
```
plot.inc.nitrate.eb
```



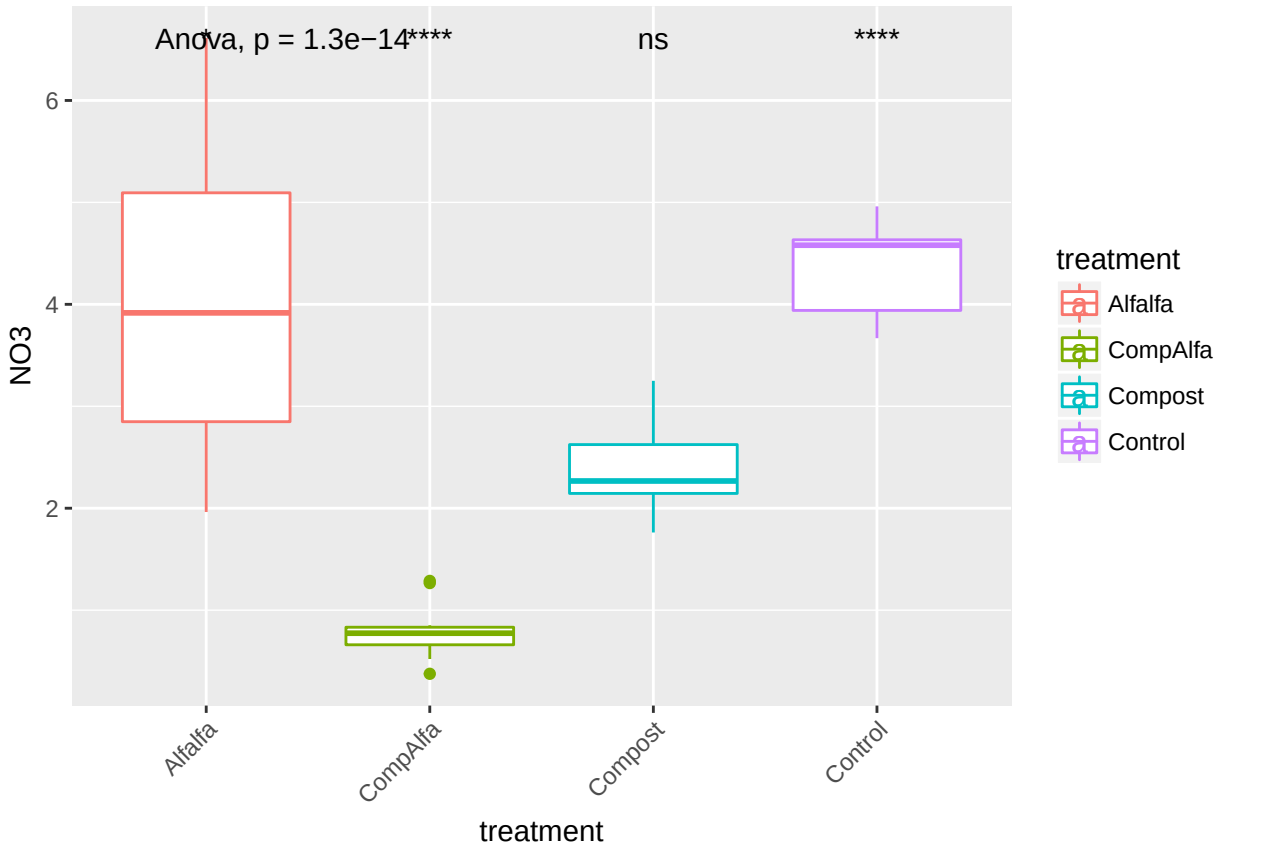

```
inc.nitrate
```


40
30
20
10
0
NO3
0
7
14
21
35
49
97
day
treatment
Alfalfa
CompAlfa
Compost
Control

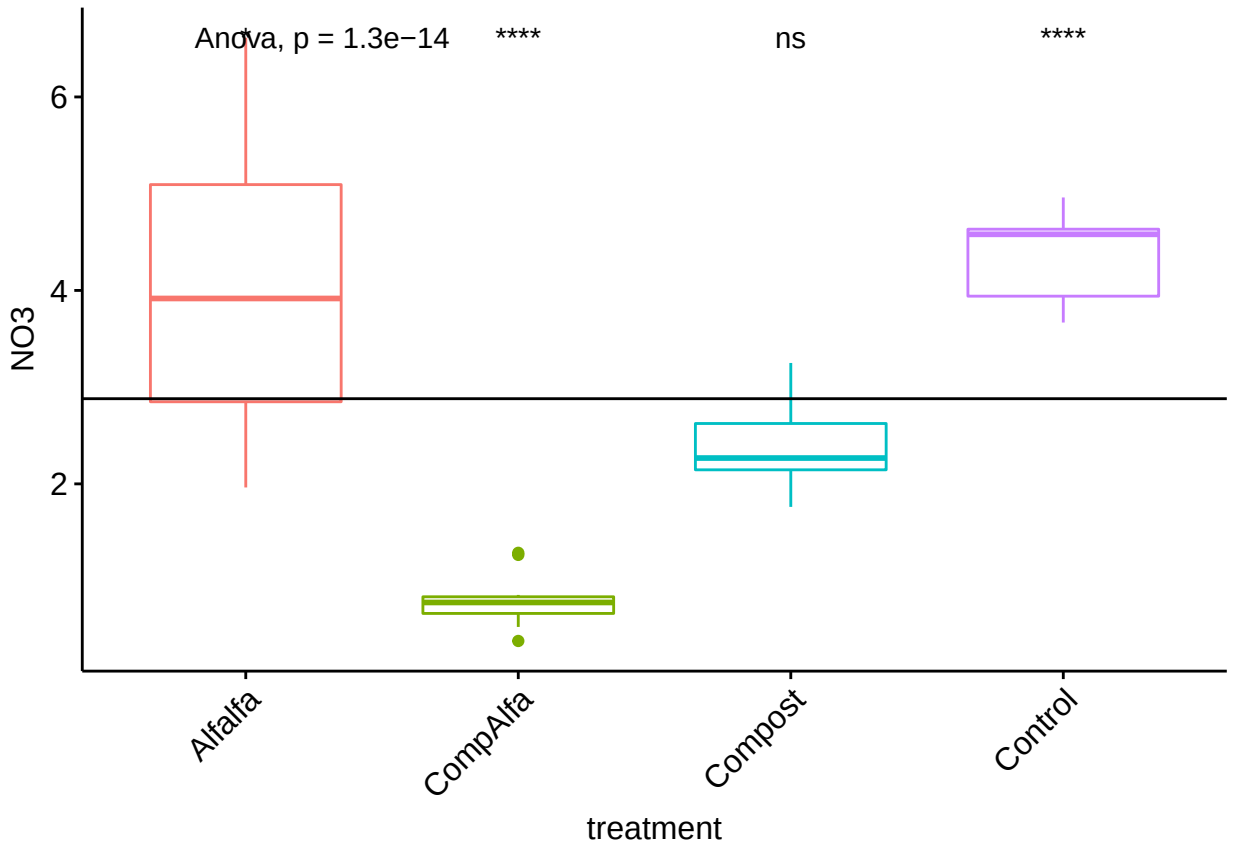
```
plot
```


0
7
14
21
35
49
97
40
30
20
10
0
NO3
treatment
Alfalfa
CompAlfa
Compost
Control

```
inc.data.7
```

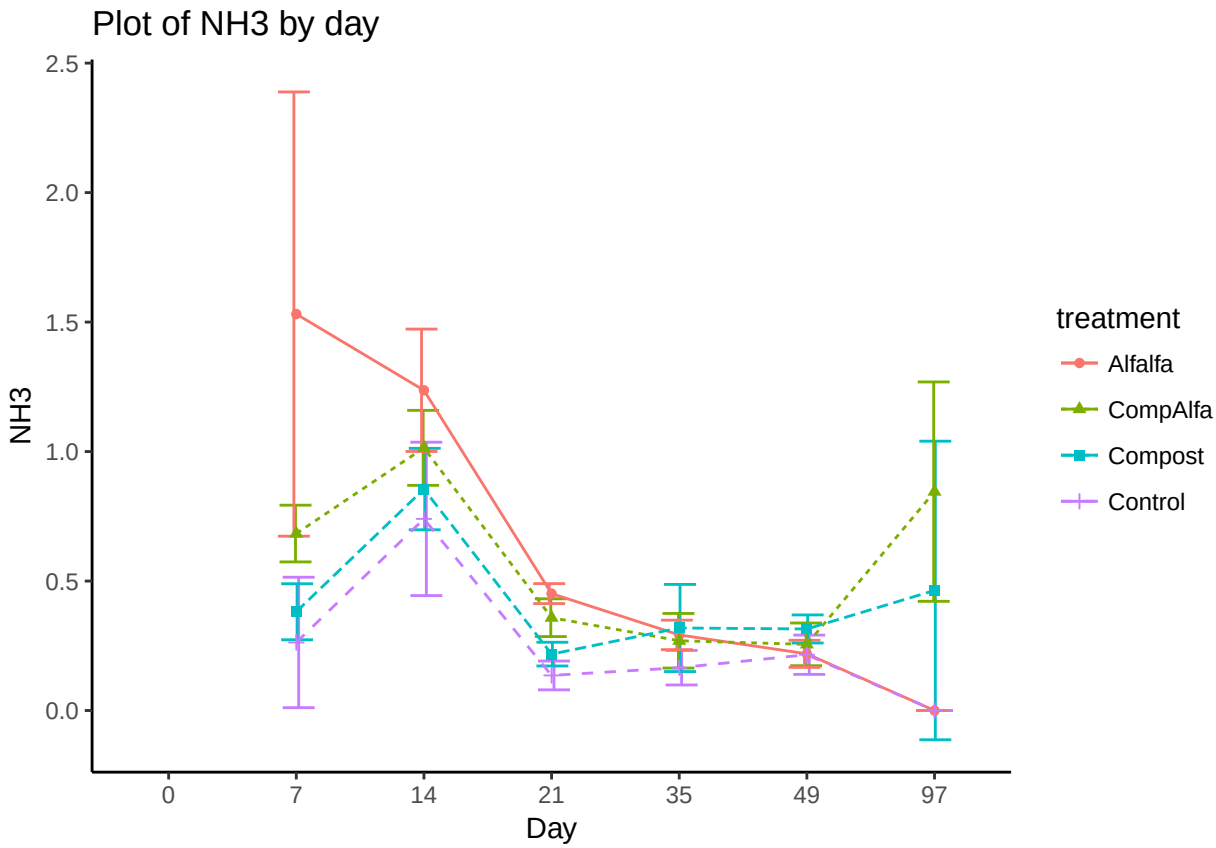

Anova, p = 1.3e−14
****
ns
****
6
4
2
NO3
Alfalfa
CompAlfa
Compost
Control
treatment
treatment
Alfalfa
CompAlfa
Compost
Control

```
stat.day.7
```

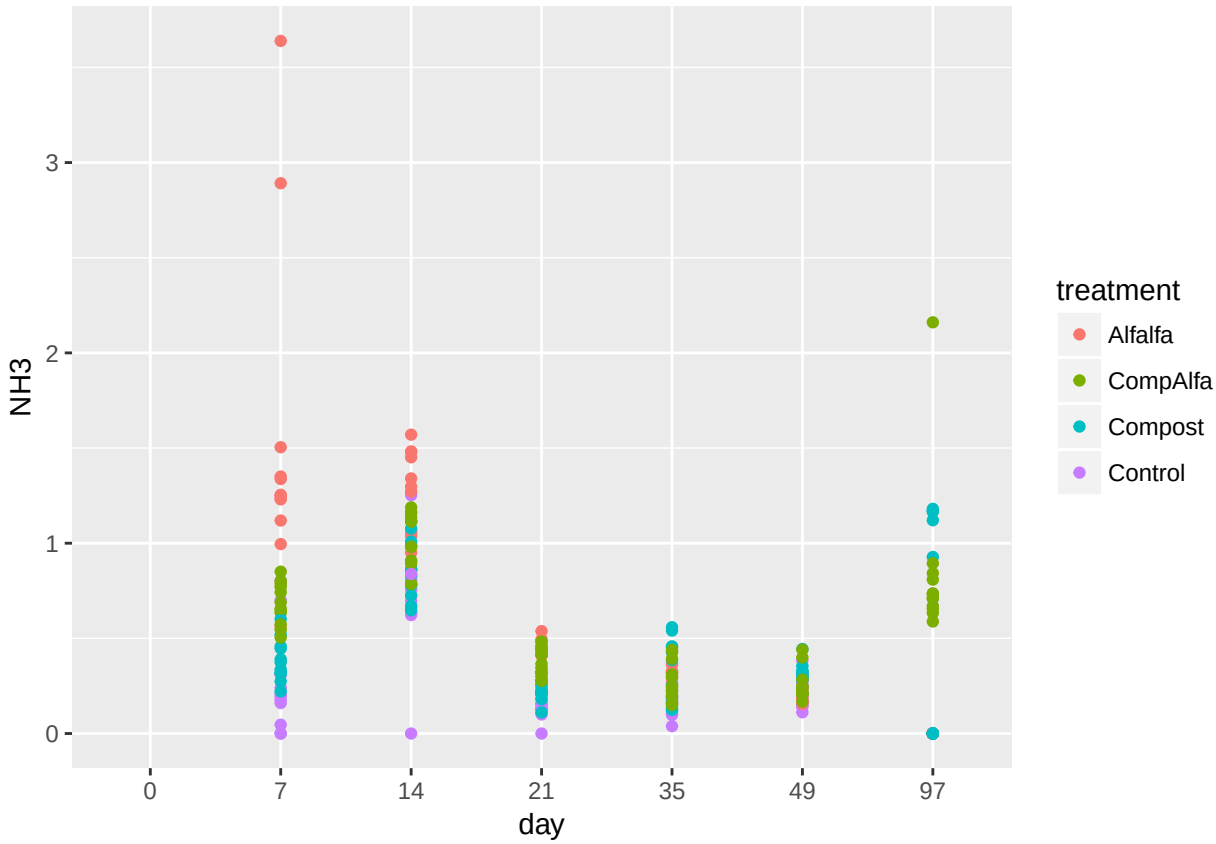


Amonia "NH3"

```
inc.amonia.eb <- DataSummary(inc.data, varname = "NH3", groupnames = c("day", "treatment"))

plot.inc.amonia.eb <- ggplot(inc.amonia.eb, aes(x = day, y = NH3, group = treatment, color = treatment))
  geom_errorbar(aes(ymin = NH3 - sd, ymax = NH3 + sd), width = 1, position = position_dodge(0.05)) +
  geom_line(aes(linetype = treatment)) +
  geom_point(aes(shape = treatment)) + labs(title = "Plot of NH3 by day", x = "Day", y = "NH3") +
  theme_classic()
plot.inc.amonia.eb
```




```
inc.amonia <- ggplot(inc.data, aes(x = day, y = NH3, group = treatment)) + geom_point(aes(color = treatı
inc.amonia
```




```
tiff('Images/inc.amonia.tiff', units="in", width=5, height=5, res=300)
#insert ggplot code
inc.amonia
dev.off()
```

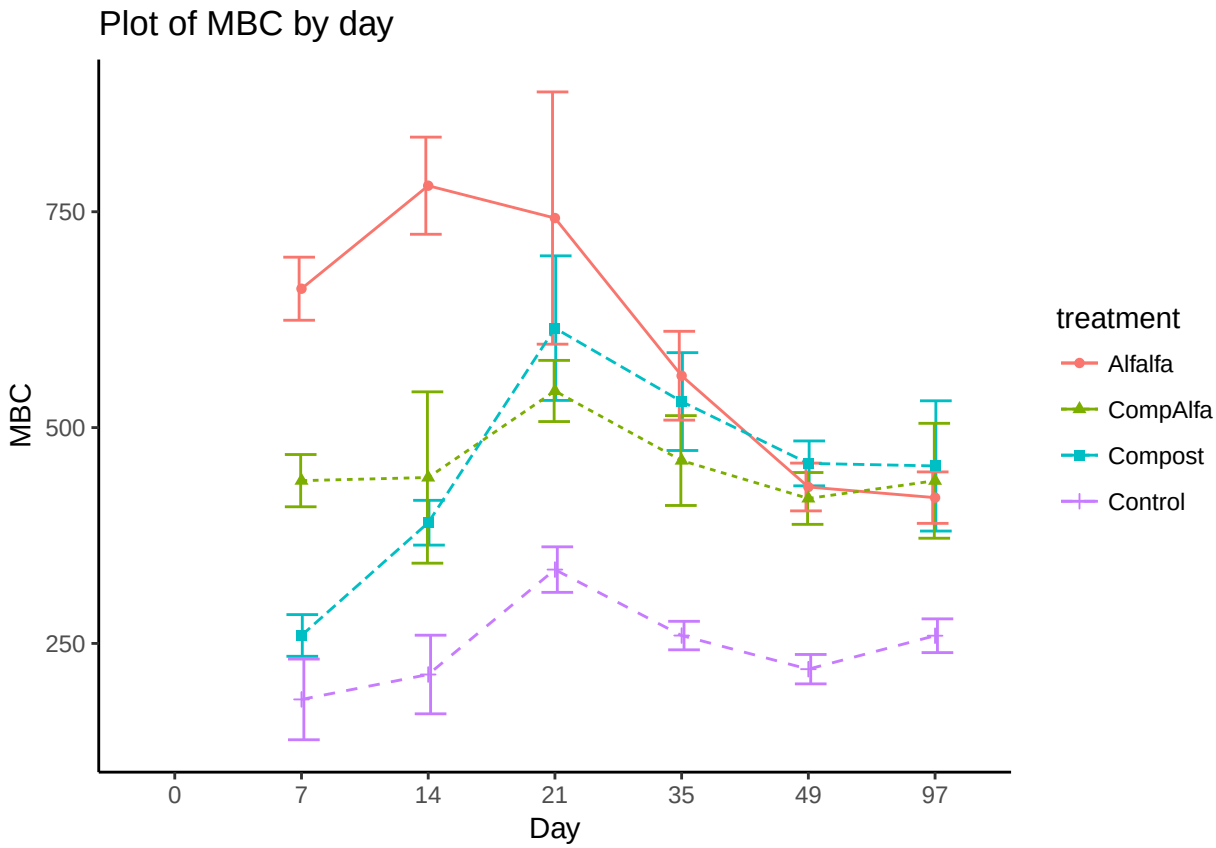
```
## pdf
##   2
```

```
tiff('Images/plot.inc.amonia.eb.tiff', units="in", width=5, height=5, res=300)
#insert ggplot code
plot.inc.amonia.eb
dev.off()
```

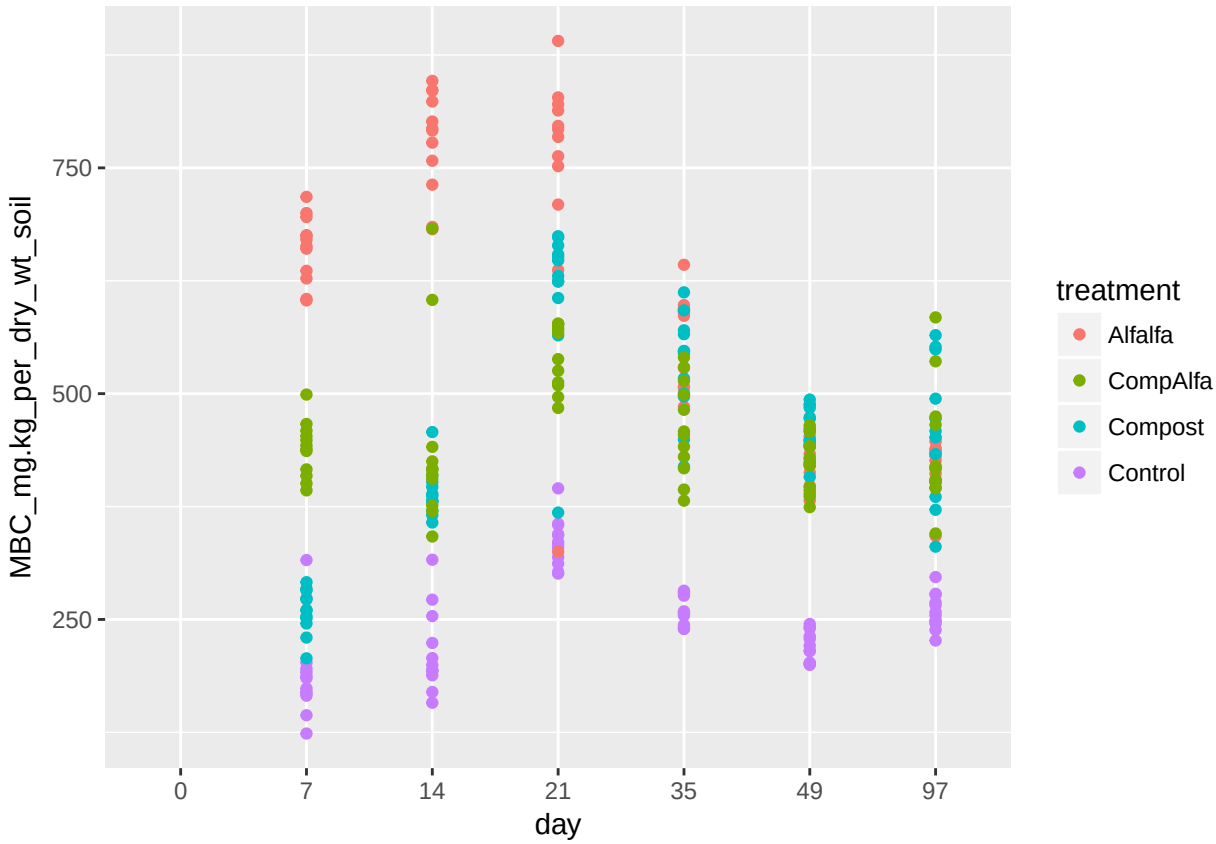
```
## pdf
##   2
```

Microbial Biomass

```
inc.micr.biomass.c.eb <- DataSummary(inc.data, varname = "MBC_mg.kg_per_dry_wt_soil", groupnames = c("da

plot.inc.micr.biomass.c.eb <- ggplot(inc.micr.biomass.c.eb, aes(x = day, y = MBC_mg.kg_per_dry_wt_soil,
  geom_errorbar(aes(ymin = MBC_mg.kg_per_dry_wt_soil - sd, ymax = MBC_mg.kg_per_dry_wt_soil + sd), width
  geom_line(aes(linetype = treatment)) +
  geom_point(aes(shape = treatment)) +
  labs(title = "Plot of MBC by day", x = "Day", y = "MBC") +
  theme_classic()
plot.inc.micr.biomass.c.eb
```




```
inc.micr.biomass.c <- ggplot(inc.data, aes(x = day, y = MBC_mg.kg_per_dry_wt_soil, group = treatment)) +
inc.micr.biomass.c
```




```
tiff('Images/plot.inc.micr.biomass.c.eb.tiff', units="in", width=5, height=5, res=300)
plot.inc.micr.biomass.c.eb
dev.off()
```

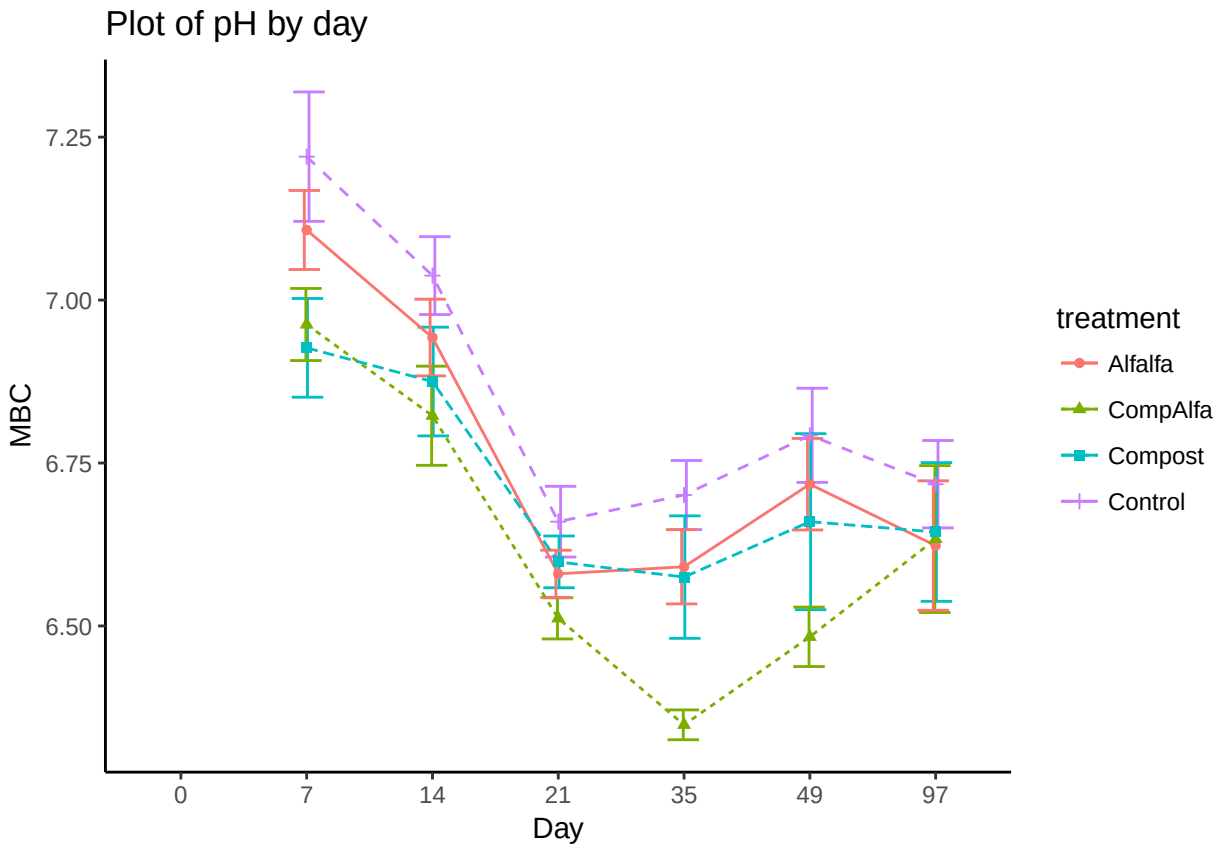
```
## pdf
##   2
```

```
tiff('Images/inc.micr.biomass.c.tiff', units="in", width=5, height=5, res=300)
inc.micr.biomass.c
dev.off()
```

```
## pdf
##   2
```

pH

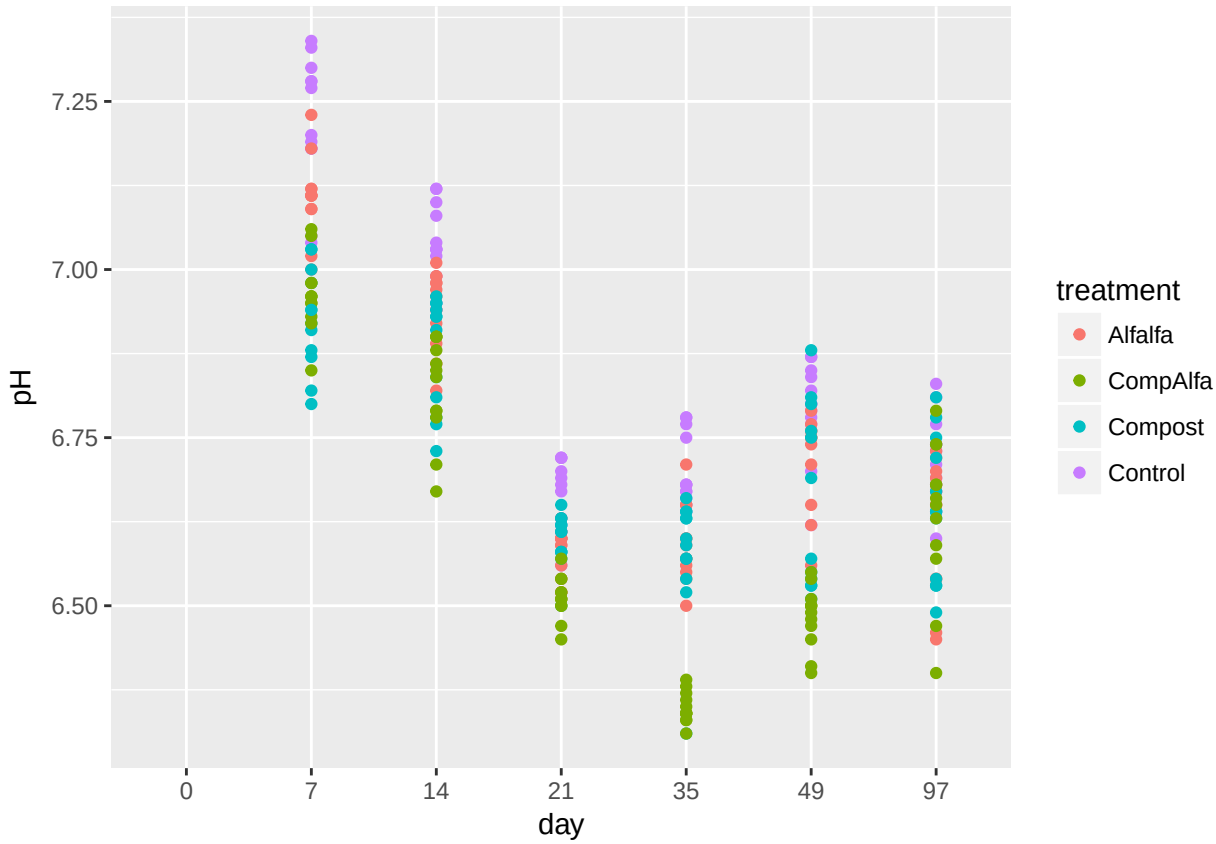
```
inc.pH.eb <- DataSummary(inc.data, varname = "pH", groupnames = c("day", "treatment"))

plot.inc.pH.eb <- ggplot(inc.pH.eb, aes(x = day, y = pH, group = treatment, color = treatment)) +
  geom_errorbar(aes(ymin = pH - sd, ymax = pH + sd), width = 1, position = position_dodge(0.05)) +
  geom_line(aes(linetype = treatment)) +
  geom_point(aes(shape = treatment)) +
  labs(title = "Plot of pH by day", x = "Day", y = "MBC") +
  theme_classic()
plot.inc.pH.eb
```




```
tiff('Images/plot.inc.pH.eb.tiff', units="in", width=5, height=5, res=300)
plot.inc.pH.eb
dev.off()
```

```
## pdf
##   2
```

```
inc.pH <- ggplot(inc.data, aes(x = day, y = pH, group = treatment)) + geom_point(aes(color = treatment))
inc.pH
```

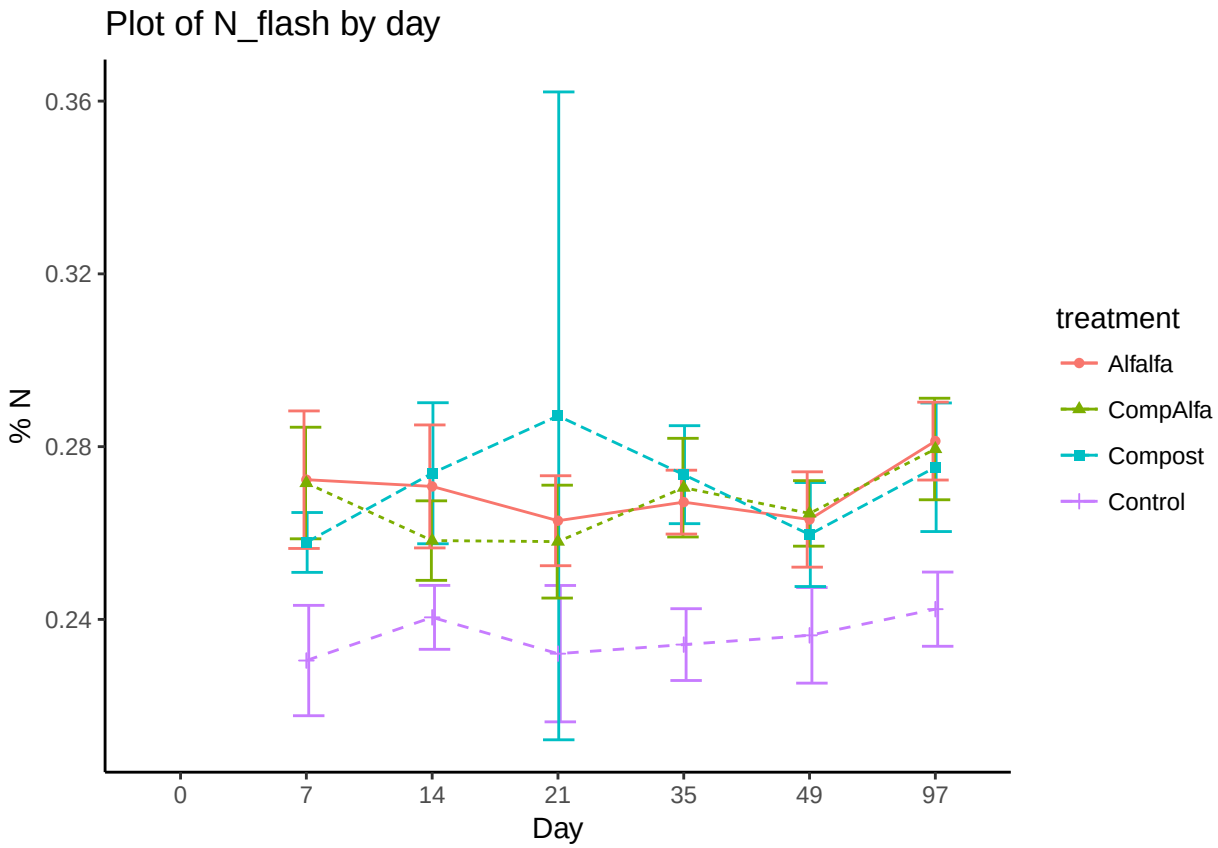



```
tiff('Images/inc.pH.tiff', units="in", width=5, height=5, res=300)
inc.pH
dev.off()
```

```
## pdf
##   2
```

Total Nitrogen via combustion analysis

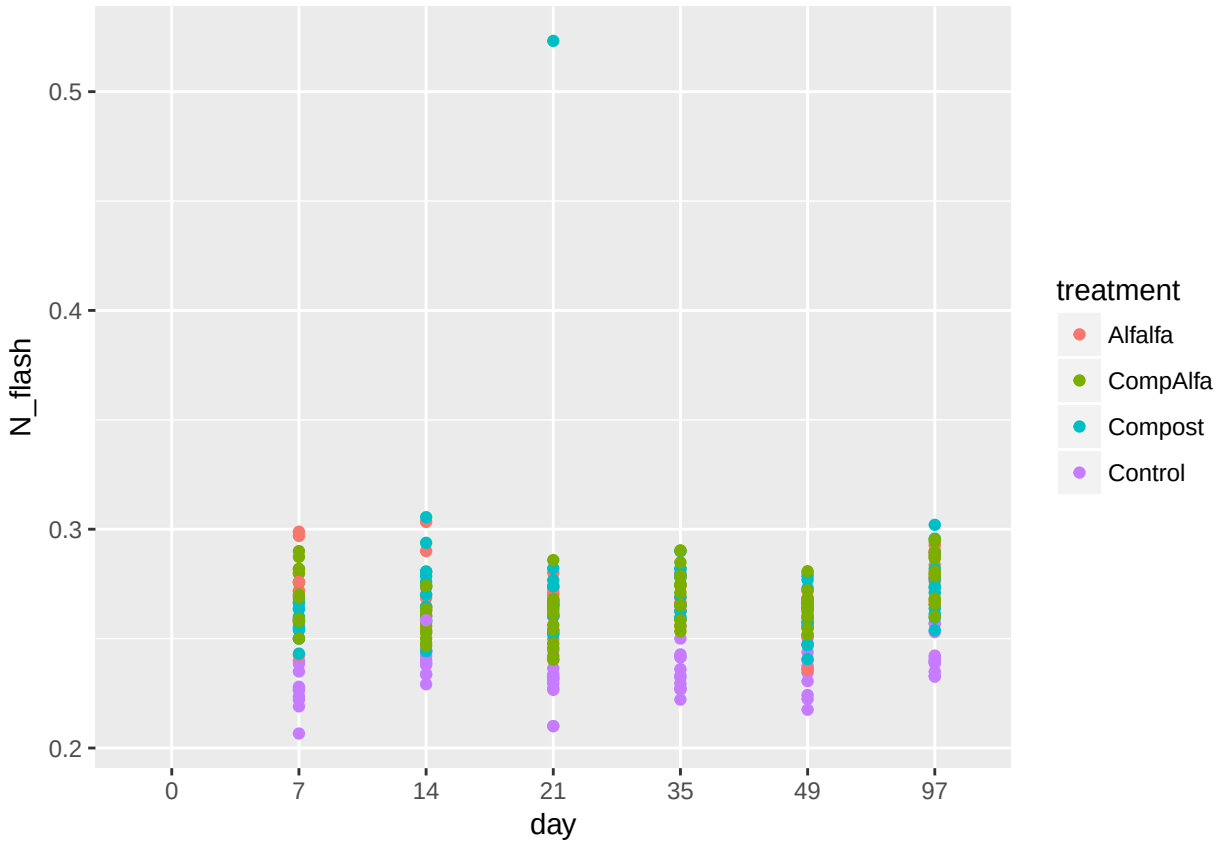
```
inc.N_flash.eb <- DataSummary(inc.data, varname = "N_flash", groupnames = c("day", "treatment"))

plot.inc.N_flash.eb <- ggplot(inc.N_flash.eb, aes(x = day, y = N_flash, group = treatment, color = trea
  geom_errorbar(aes(ymin = N_flash - sd, ymax = N_flash + sd), width = 1, position = position_dodge(0.0
  geom_line(aes(linetype = treatment)) +
  geom_point(aes(shape = treatment)) +
  labs(title = "Plot of N_flash by day", x = "Day", y = "% N") +
  theme_classic()
plot.inc.N_flash.eb
```




```
tiff('Images/plot.inc.N_flash.eb.tiff', units="in", width=5, height=5, res=300)
plot.inc.N_flash.eb
dev.off()
```

```
## pdf
##   2
```

```
inc.N_flash <- ggplot(inc.data, aes(x = day, y = N_flash, group = treatment)) + geom_point(aes(color = 
inc.N_flash
```

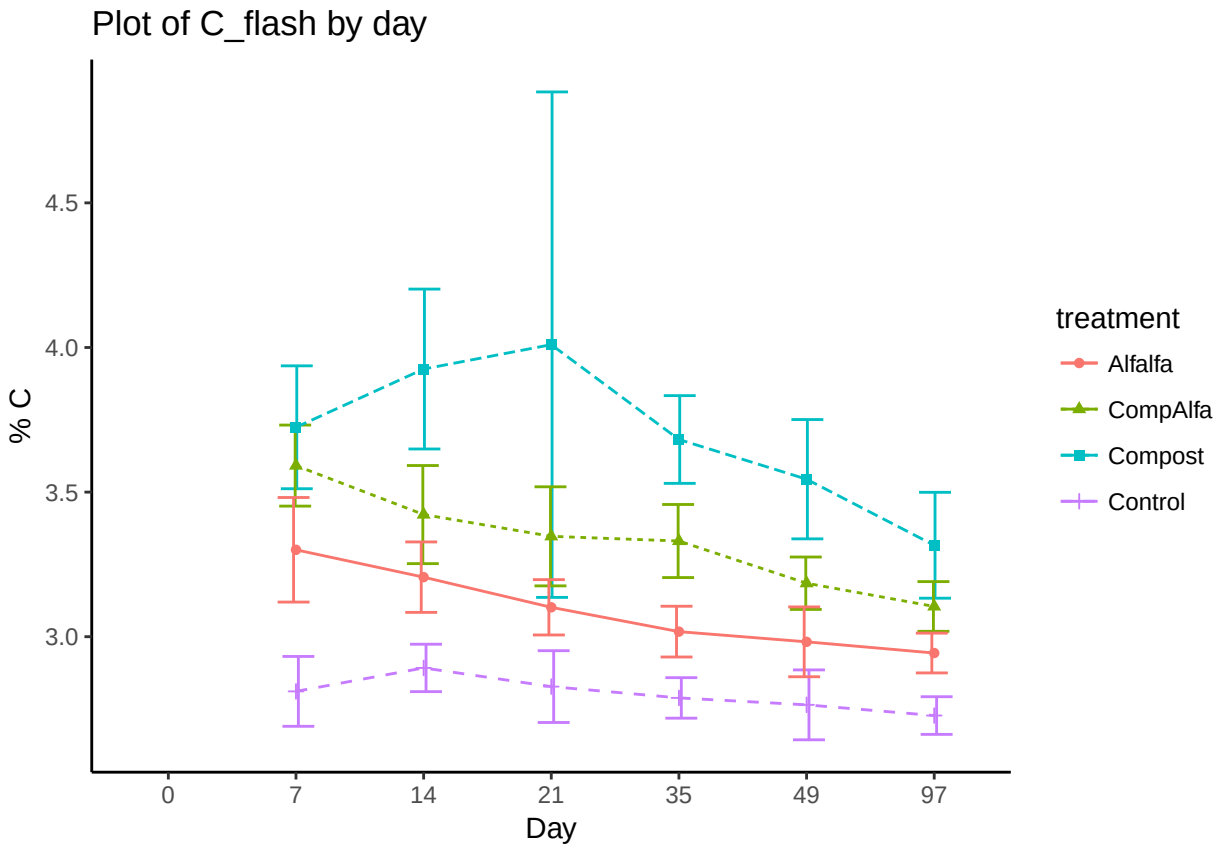



```
tiff('Images/inc.N_flash.tiff', units="in", width=5, height=5, res=300)
inc.N_flash
dev.off()
```

```
## pdf
##   2
```

Total Carbon via combustion analysis

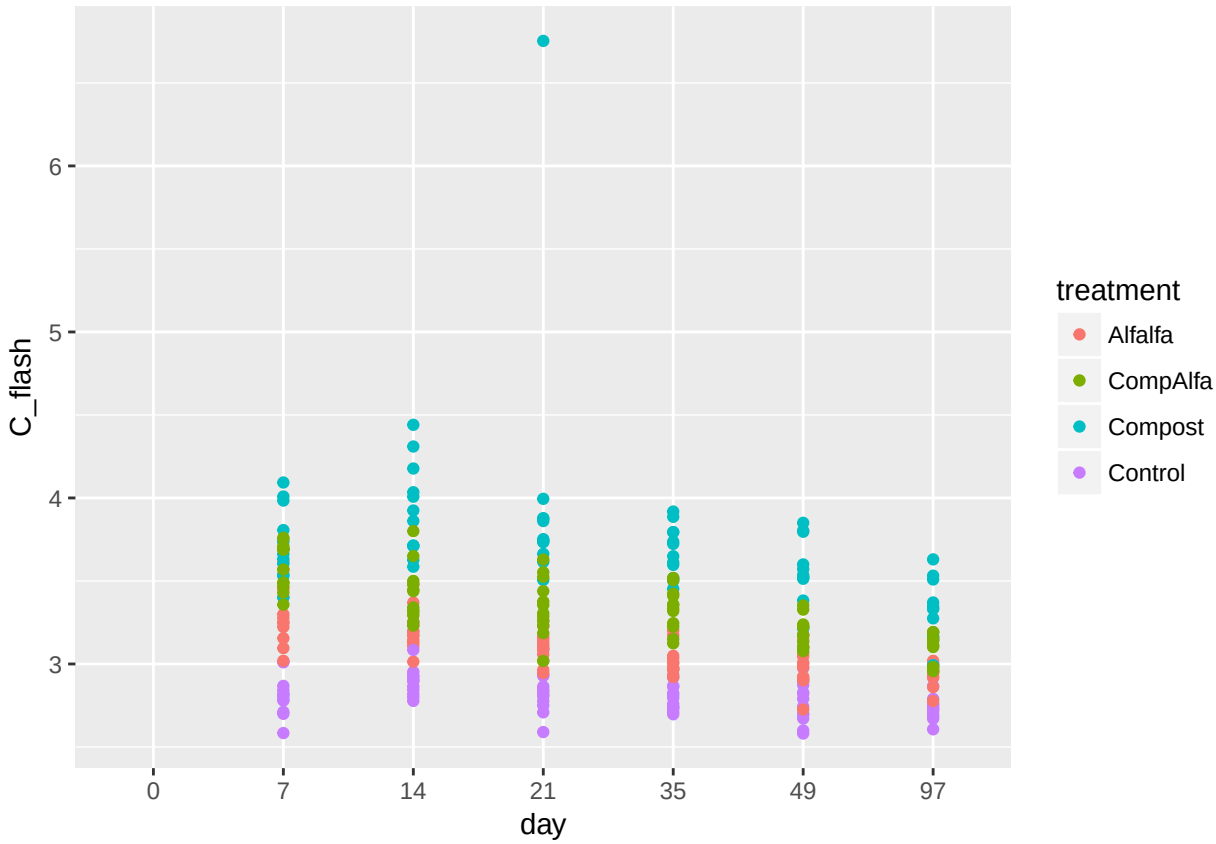
```
inc.C_flash.eb <- DataSummary(inc.data, varname = "C_flash", groupnames = c("day", "treatment"))

plot.inc.C_flash.eb <- ggplot(inc.C_flash.eb, aes(x = day, y = C_flash, group = treatment, color = trea
  geom_errorbar(aes(ymin = C_flash - sd, ymax = C_flash + sd), width = 1, position = position_dodge(0.0
  geom_line(aes(linetype = treatment)) +
  geom_point(aes(shape = treatment)) +
  labs(title = "Plot of C_flash by day", x = "Day", y = "% C") +
  theme_classic()
plot.inc.C_flash.eb
```




```
tiff('Images/inc.C_flash.tiff', units="in", width=5, height=5, res=300)
plot.inc.C_flash.eb
dev.off()
```

```
## pdf
##   2
```

```
inc.C_flash <- ggplot(inc.data, aes(x = day, y = C_flash, group = treatment)) + geom_point(aes(color = 
inc.C_flash
```

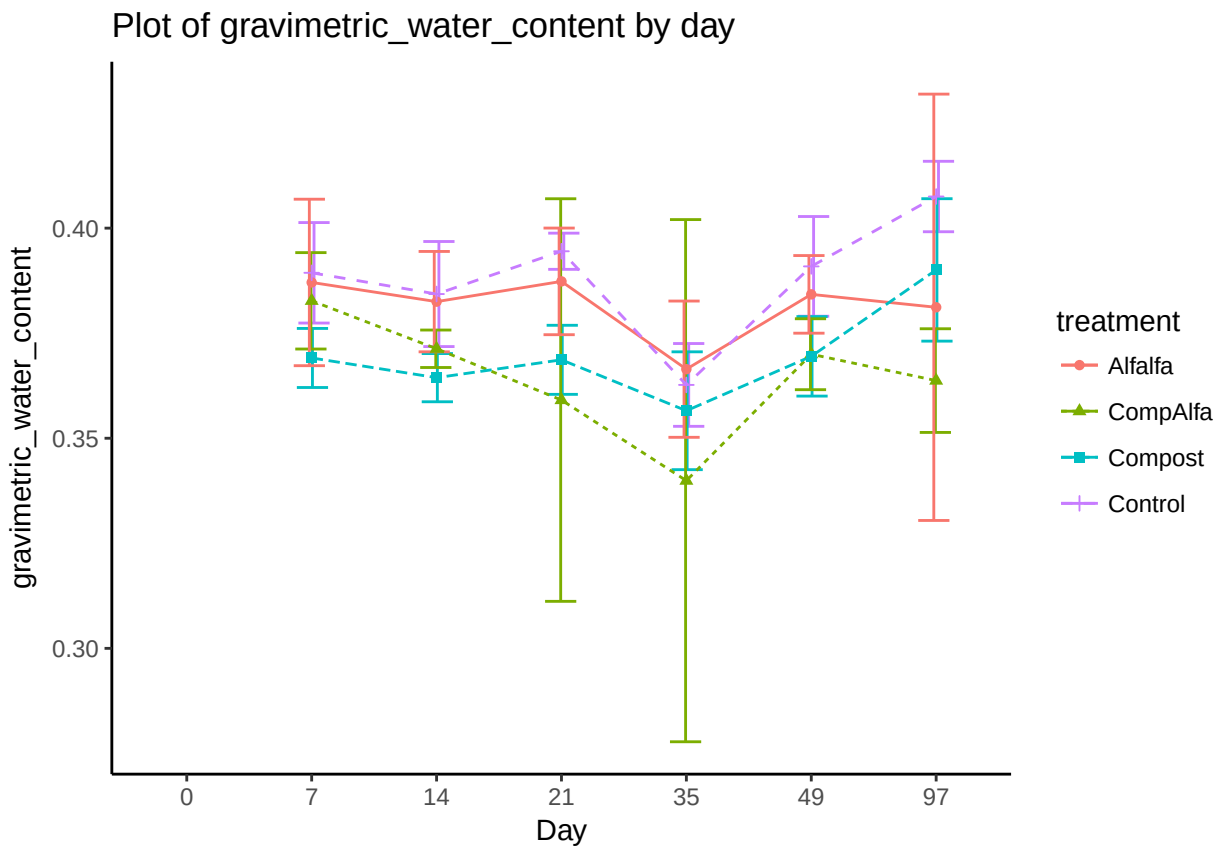



```
tiff('Images/inc.C_flash.tiff', units="in", width=5, height=5, res=300)
inc.C_flash
dev.off()
```

```
## pdf
##   2
```

Gravimetric moisture content

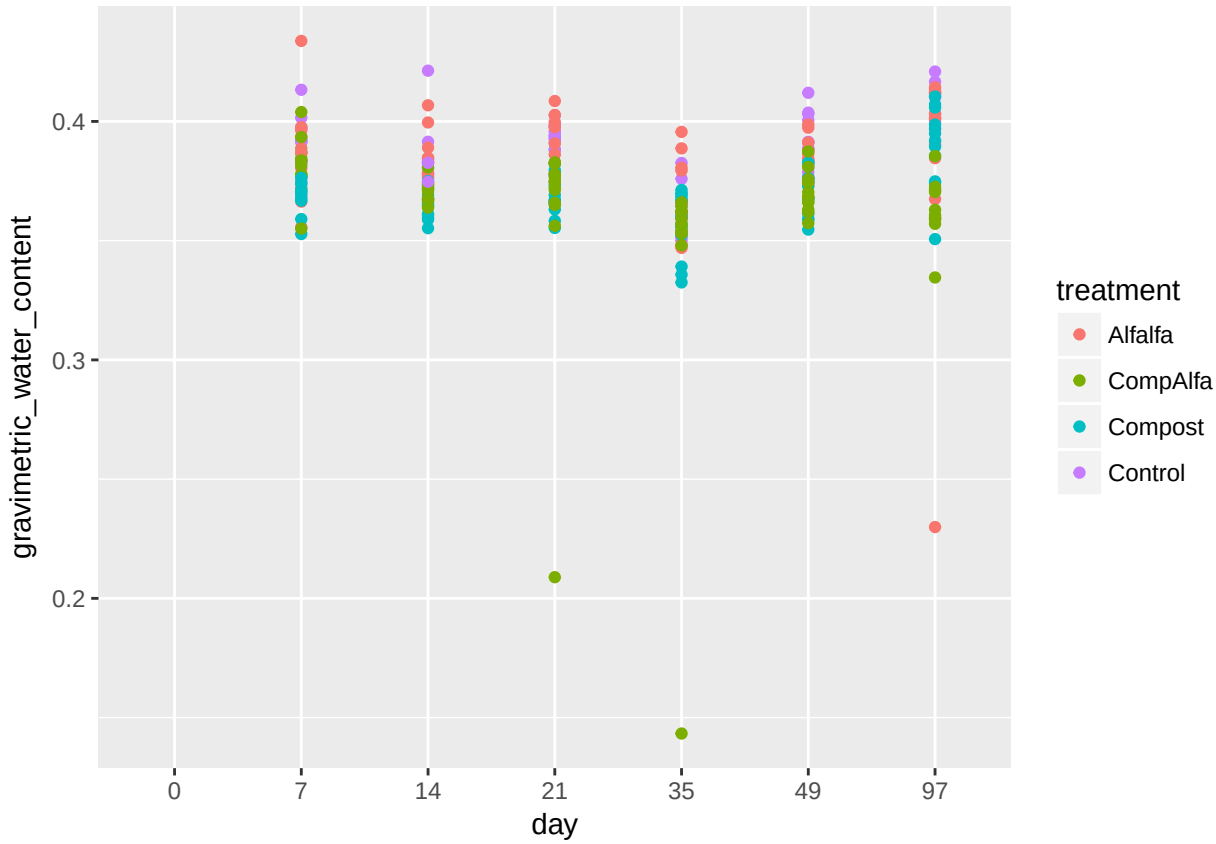
```
inc.gravimetric_water_content.eb <- DataSummary(inc.data, varname = "gravimetric_water_content", groupna

plot.inc.gravimetric_water_content.eb <- ggplot(inc.gravimetric_water_content.eb, aes(x = day, y = gravi
  geom_errorbar(aes(ymin = gravimetric_water_content - sd, ymax = gravimetric_water_content + sd), width
  geom_line(aes(linetype = treatment)) +
  geom_point(aes(shape = treatment)) +
  labs(title = "Plot of gravimetric_water_content by day", x = "Day", y = "gravimetric_water_content") +
  theme_classic()
plot.inc.gravimetric_water_content.eb
```




```
tiff('Images/plot.inc.gravimetric_water_content.eb.tiff', units="in", width=5, height=5, res=300)
plot.inc.gravimetric_water_content.eb
dev.off()
```

```
## pdf
##   2
```

```
inc.gravimetric_water_content <- ggplot(inc.data, aes(x = day, y = gravimetric_water_content, group = t
inc.gravimetric_water_content
```




```
tiff('Images/inc.gravimetric_water_content.tiff', units="in", width=5, height=5, res=300)
inc.gravimetric_water_content
dev.off()
```

```
## pdf
##   2
```

Under Construction

Try tax_glom to identify the OTUs on one group but not another

```
inc.raw.control <- subset_samples(inc.raw, treatment == "Control" & day == "0")
inc.raw.alfalfa <- subset_samples(inc.raw, treatment == "Alfalfa" & day == "0")
# Day zero comparison of control and alfalfa OTUs
control.no.0 <- filter_taxa(inc.raw.control, function(x) sum(x) >0, TRUE)
alfalfa.no.0 <- filter_taxa(inc.raw.alfalfa, function(x) sum(x) >0, TRUE)
control.taxa <- rownames(tax_table(control.no.0))
alfalfa.taxa <- rownames(tax_table(alfalfa.no.0))
length(intersect(control.taxa, alfalfa.taxa))
```

```
## [1] 3284
```

```
# OTUs in alfalfa day 0 only
only.alfalfa <- setdiff(alfalfa.taxa, control.taxa)
length(only.alfalfa)
```

```
## [1] 1573
```

```
tax.in.alf <- tax_table(inc.raw.alfalfa)[only.alfalfa]
# Below melt for plotting and prune to get taxa from larger phyloseq object
#only.alfalfa.day.0 <- prune_taxa(only.alfalfa, inc.raw.alfalfa)
#only.alfalfa.day.0.df <- psmelt(only.alfalfa.day.0)
```