TN - SH Project - Williams

Skye Wills

**TN soil Health project - KSSL Data, Data Analysis**

## R SetUp

This setup 'chunk' loads needed This portion may need to be modified with updated info for your file locations - alter path inside quotations, note that the backslash is used instead of the forward slash. This ensures that input and output files come from the correct locations

## Required packages

knitr::opts\_chunk$set(echo = FALSE, comment = "#", warning = FALSE, message = FALSE, error =FALSE, tidy.opts=list(width.cutoff=60),tidy=TRUE)  
  
rm(list=ls()) #clear previous data  
   
list.of.packages <- c("knitr", "fastmatch", "ggplot2", "Rcpp", "RColorBrewer", "lattice", "maps", "aqp", "soilDB", "tidyr", "tidyverse", "ggthemes", "ggmosaic", "RCurl", "ggjoy", "broom", "lme4", "ggjoy", "ggbeeswarm", "viridis", 'cetcolor')  
  
   
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]  
if(length(new.packages)) install.packages(new.packages)  
   
library(cetcolor)

## Warning: package 'cetcolor' was built under R version 3.4.2

library(knitr)

## Warning: package 'knitr' was built under R version 3.4.3

library(fastmatch)

## Warning: package 'fastmatch' was built under R version 3.4.1

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.1

library(Rcpp)

## Warning: package 'Rcpp' was built under R version 3.4.1

library(RColorBrewer)  
library(lattice)  
library(maps)

## Warning: package 'maps' was built under R version 3.4.3

library(aqp)  
library(soilDB)  
library(stringr)

## Warning: package 'stringr' was built under R version 3.4.1

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.4.3

## Warning: package 'tibble' was built under R version 3.4.3

## Warning: package 'tidyr' was built under R version 3.4.3

## Warning: package 'readr' was built under R version 3.4.1

## Warning: package 'purrr' was built under R version 3.4.3

## Warning: package 'dplyr' was built under R version 3.4.2

## Warning: package 'forcats' was built under R version 3.4.1

library(tidyr)  
library(ggthemes)

## Warning: package 'ggthemes' was built under R version 3.4.1

library(ggmosaic)

## Warning: package 'ggmosaic' was built under R version 3.4.1

## Warning: package 'productplots' was built under R version 3.4.1

library(RCurl)  
library(broom)

## Warning: package 'broom' was built under R version 3.4.3

library(lme4)

## Warning: package 'lme4' was built under R version 3.4.1

library(ggjoy)

## Warning: package 'ggjoy' was built under R version 3.4.1

library(ggbeeswarm)

## Warning: package 'ggbeeswarm' was built under R version 3.4.2

library(viridis)

## Warning: package 'viridis' was built under R version 3.4.3

## Warning: package 'viridisLite' was built under R version 3.4.3

#data storge and output  
#alter path as needed  
  
sh\_TN <- "D:/Disk 2/Projects/Candiss/TN\_sh"  
  
#create subfolders to save output  
Rout = paste0(sh\_TN, "/Rout/")  
ifelse(!dir.exists(file.path(Rout)), dir.create(Rout), FALSE)

## [1] FALSE

## ##Inputs

Then relabels some columns with simpler labels.

**ADD links to data**

#### Data prep

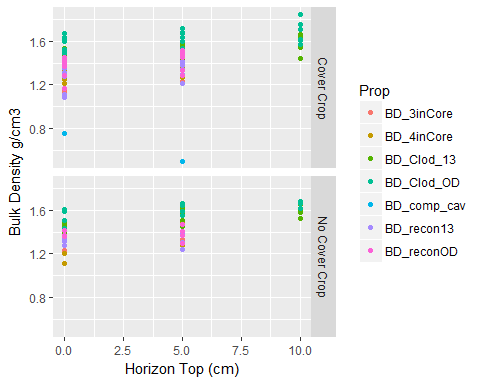
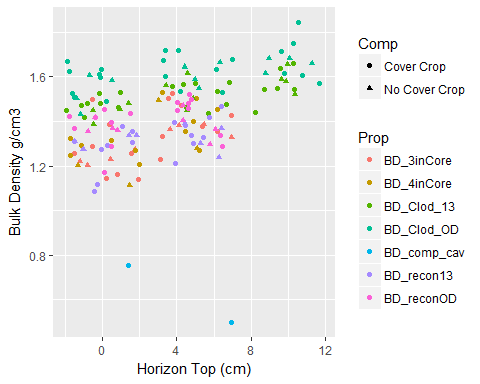
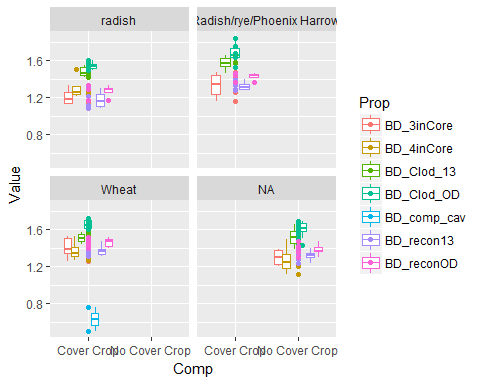
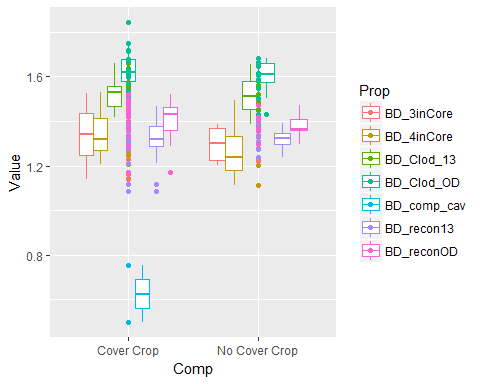
TN\_labels file has been used to rename columns from LIMS with simple labels.

Use dplyr to filter data so that only columns that we want to analyze are kept - the extras are currently labeled with X - then they are reorganized into

#### Bulk Density

Evaluate bulk density to inform combining methods in dataprep phase

# # A tibble: 7 x 3  
# Prop `Cover Crop` `No Cover Crop`  
# \* <chr> <int> <int>  
# 1 BD\_3inCore 16 8  
# 2 BD\_4inCore 12 4  
# 3 BD\_Clod\_13 22 12  
# 4 BD\_Clod\_OD 22 12  
# 5 BD\_comp\_cav 2 0  
# 6 BD\_recon13 16 8  
# 7 BD\_reconOD 16 8



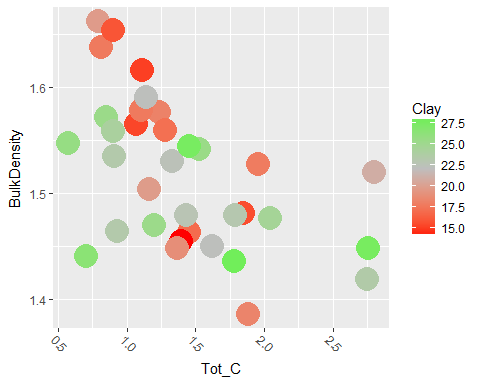
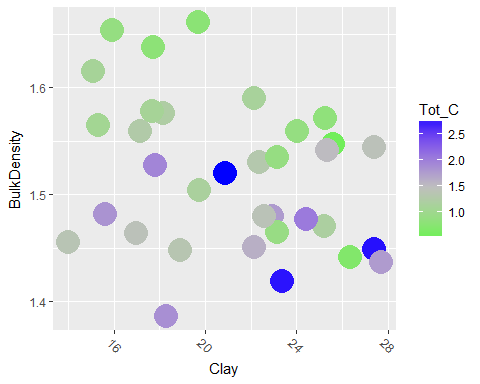
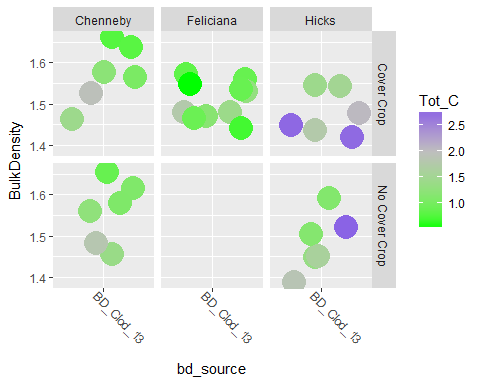
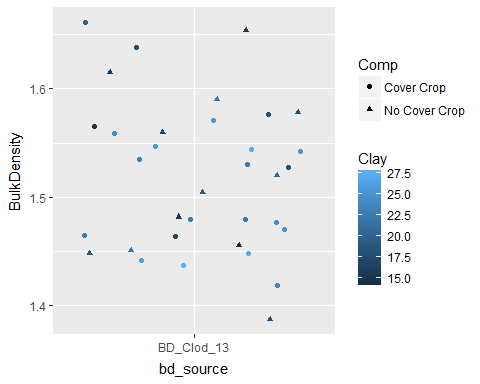
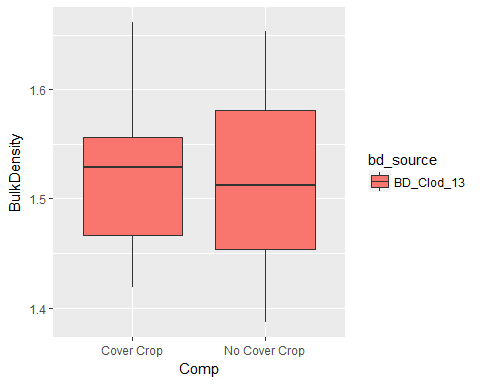
#### Db Data Prep

Input data.  
Create standardized properties when multiple methods are used, such as bulk density -change rank of methods to alter the way multiple methods are analyzed -Currently, 2015 samples use compliant cavity because it was done on all samples (**really a core**). For 2014, core bulk density was favored because it was done on nearly all samples

Add comparable layers to allow for

#   
# BD\_Clod\_13   
# 34

# Min. 1st Qu. Median Mean 3rd Qu. Max.   
# 1.386 1.464 1.523 1.518 1.564 1.662



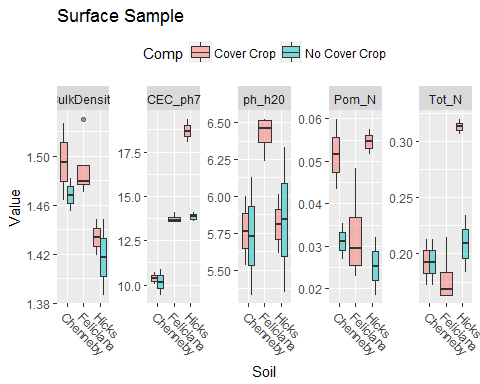
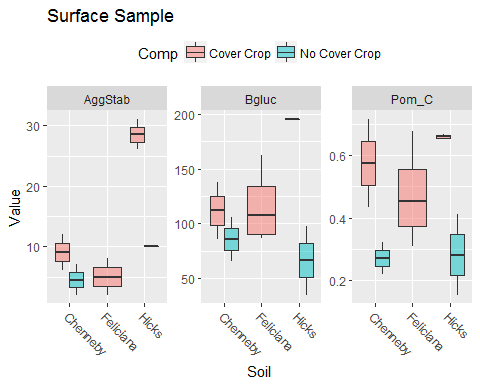
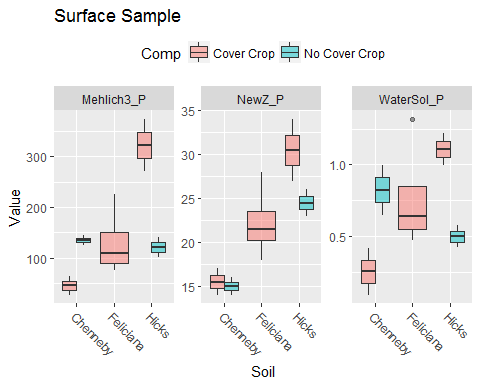
# Min. 1st Qu. Median Mean 3rd Qu. Max.   
# -0.01210 0.00000 0.03010 0.08093 0.11166 0.40039

# Min. 1st Qu. Median Mean 3rd Qu. Max.   
# 0.5713 0.9626 1.2932 1.3972 1.7076 2.7534

Removed this section - didn't include coarse fragments

### Exploratory plots

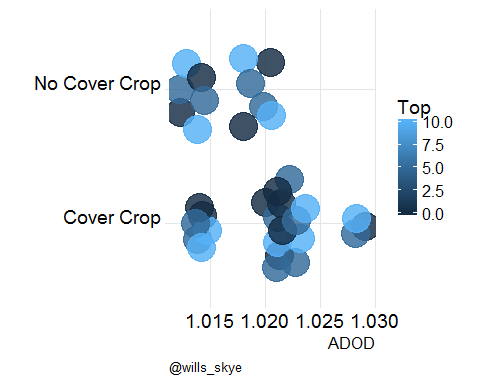
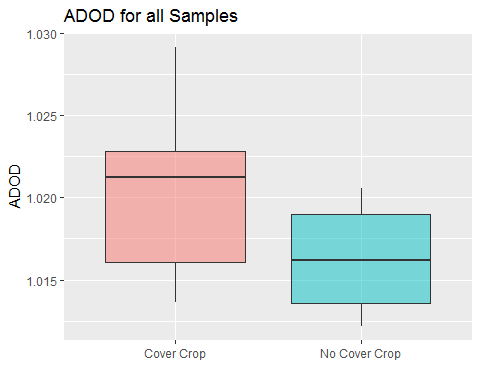
#### Surface Properties



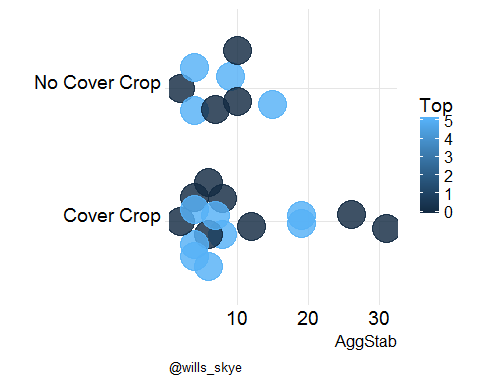
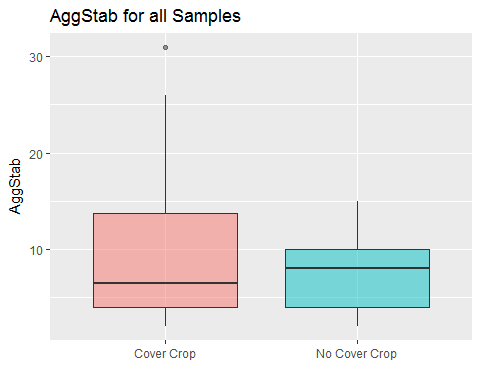
## Exploratory Plots

Individual

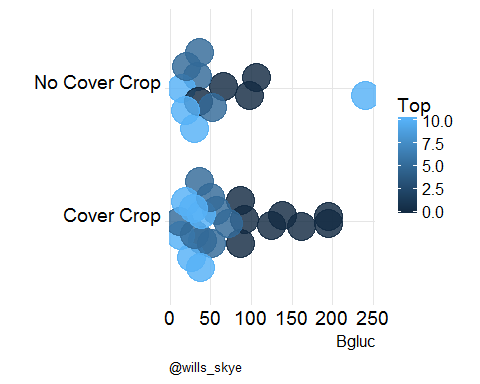
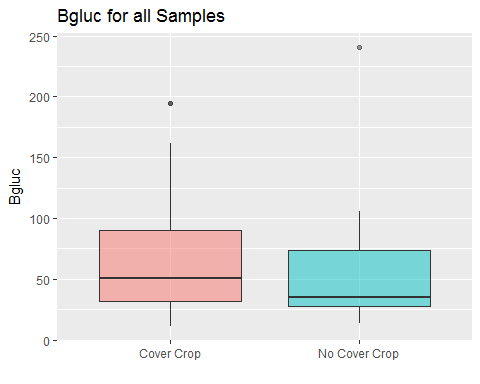
# [1] "ADOD"



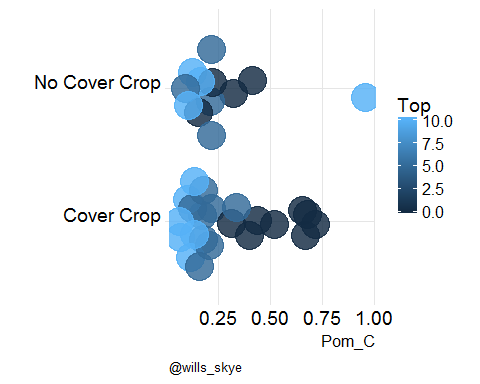
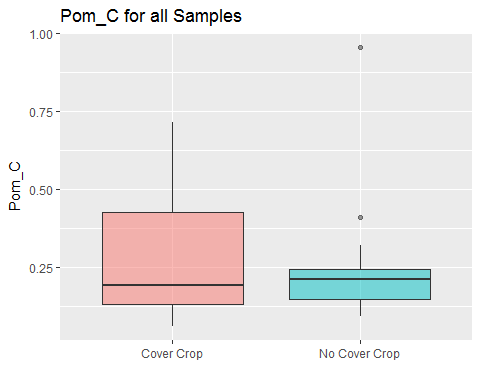
# [1] "AggStab"



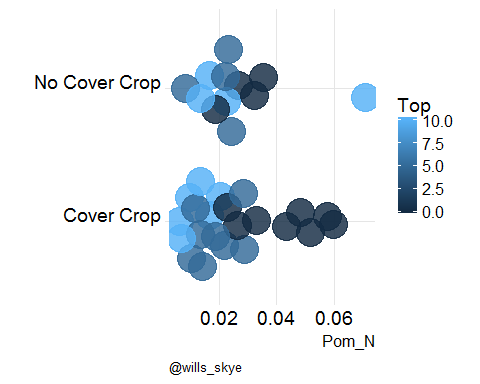
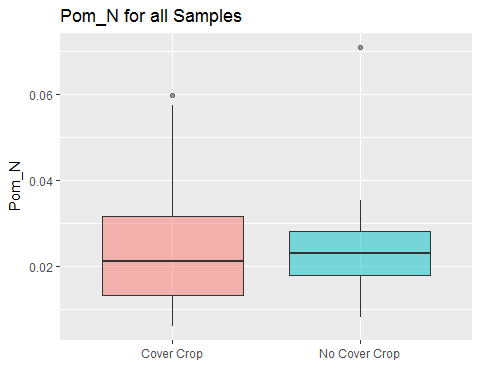
# [1] "Bgluc"



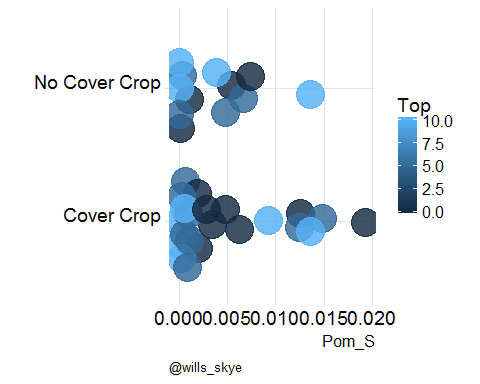
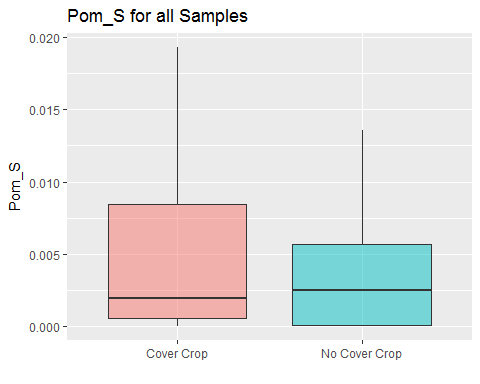
# [1] "Pom\_C"



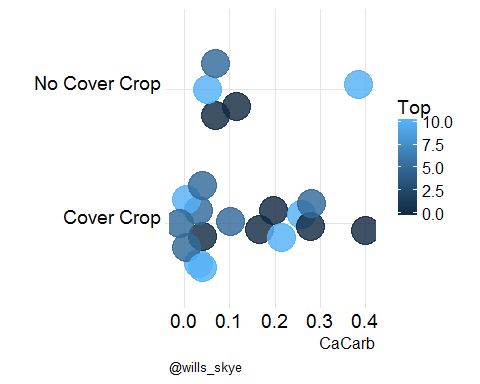
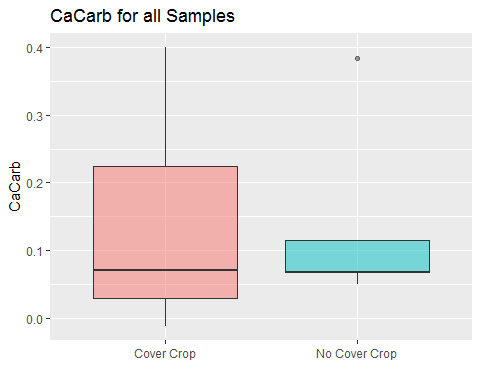
# [1] "Pom\_N"



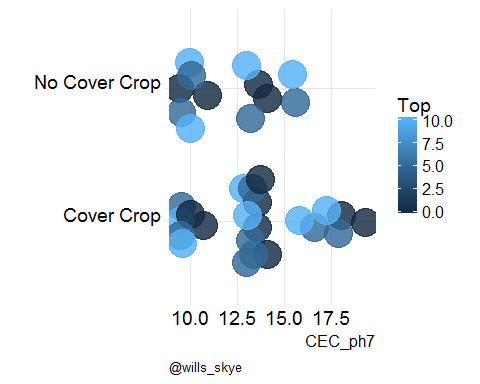
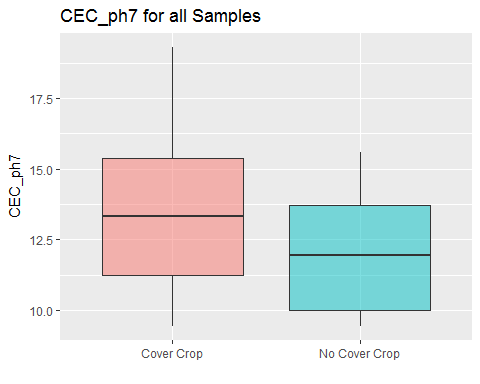
# [1] "Pom\_S"



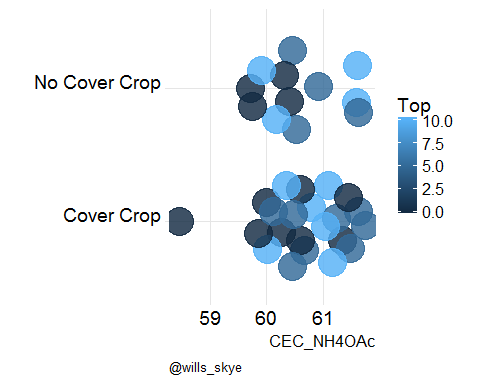
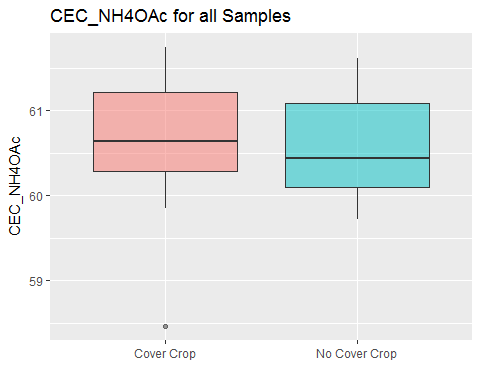
# [1] "CaCarb"



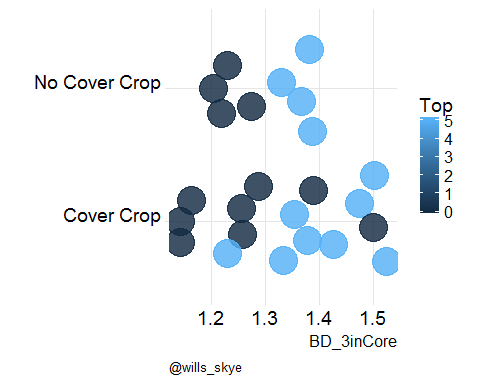
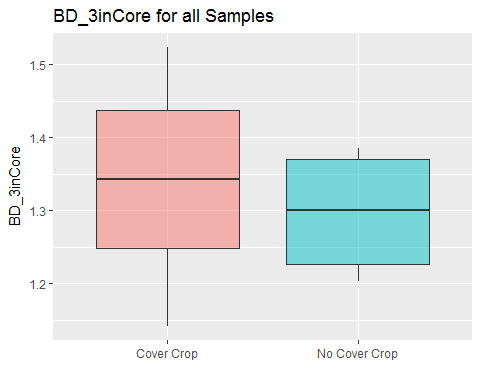
# [1] "CEC\_ph7"



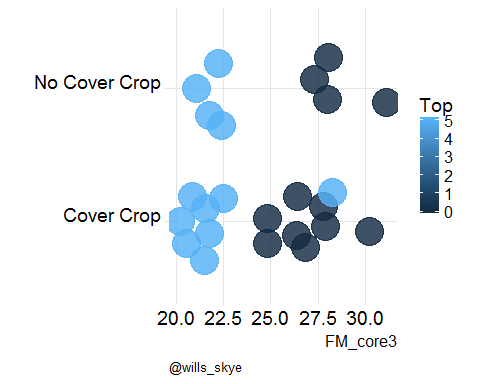
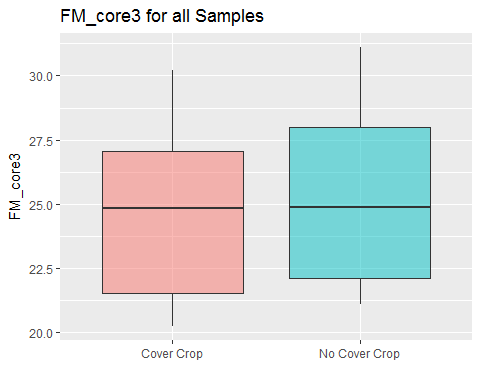
# [1] "CEC\_NH4OAc"



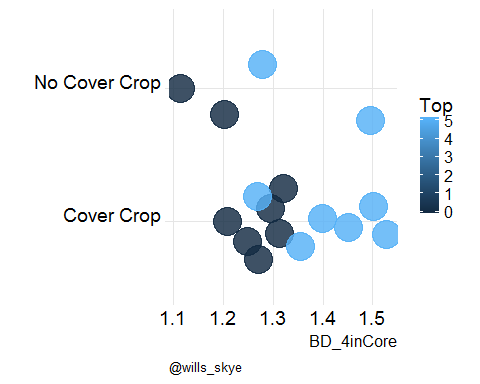
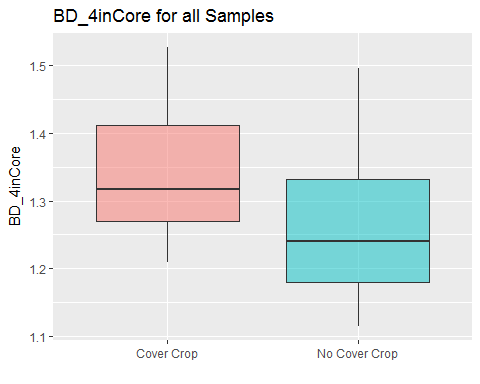
# [1] "BD\_3inCore"



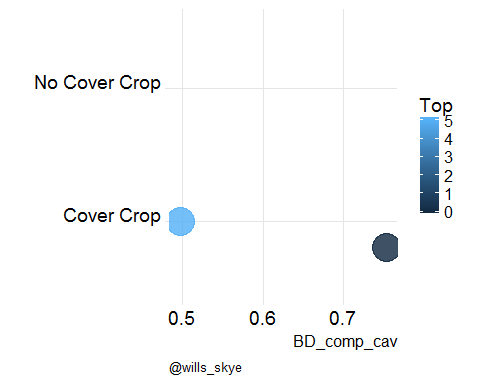
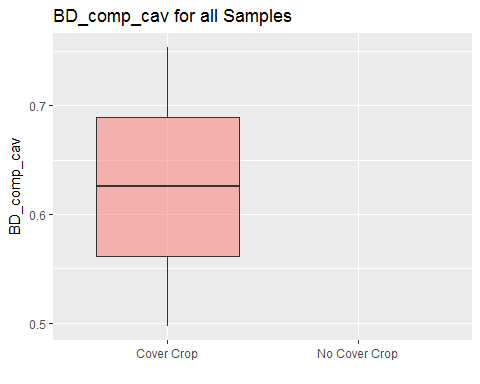
# [1] "FM\_core3"



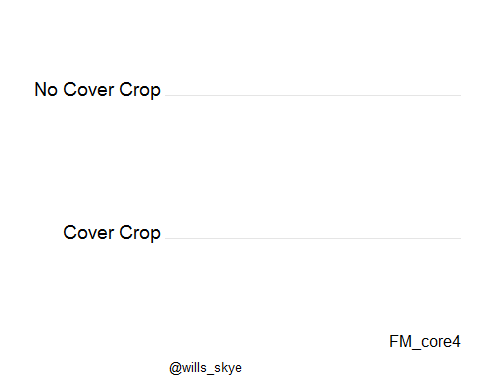
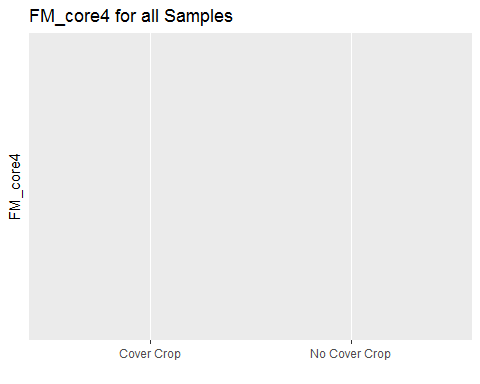
# [1] "BD\_4inCore"



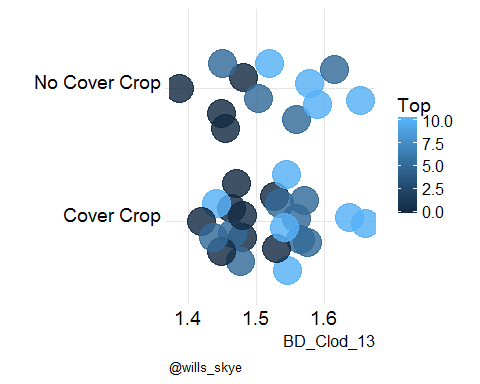
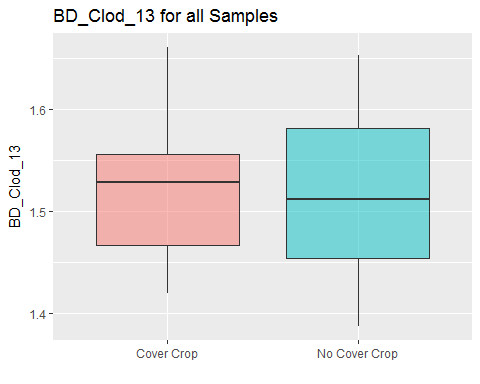
# [1] "BD\_comp\_cav"



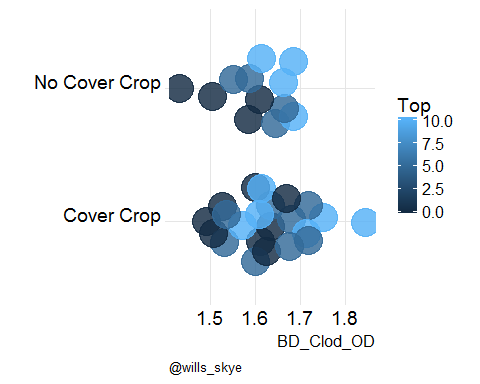
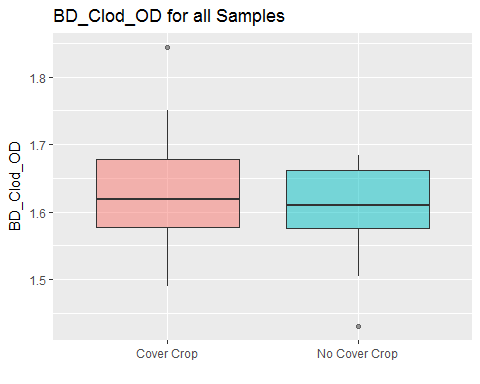
# [1] "FM\_core4"



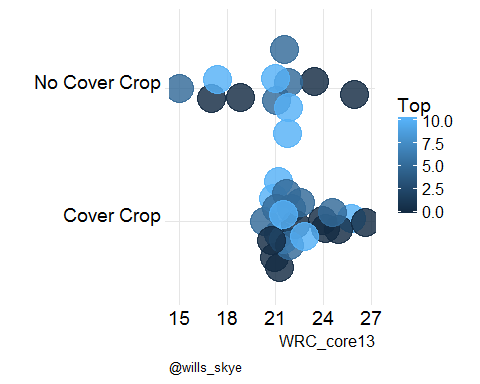
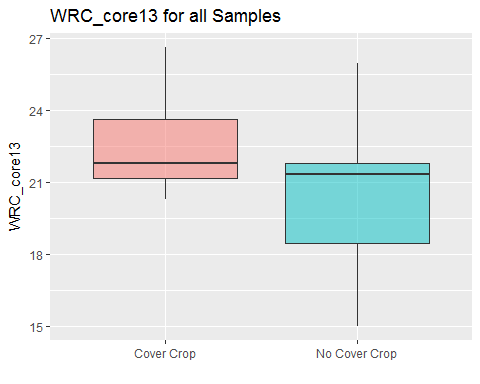
# [1] "BD\_Clod\_13"



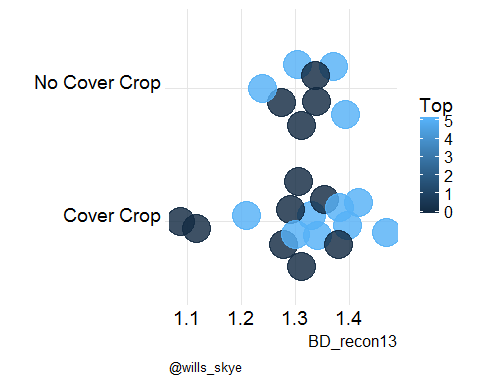
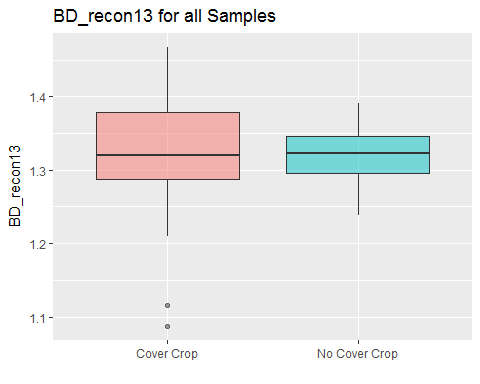
# [1] "BD\_Clod\_OD"



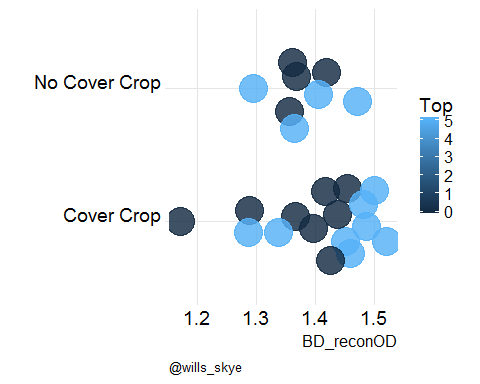
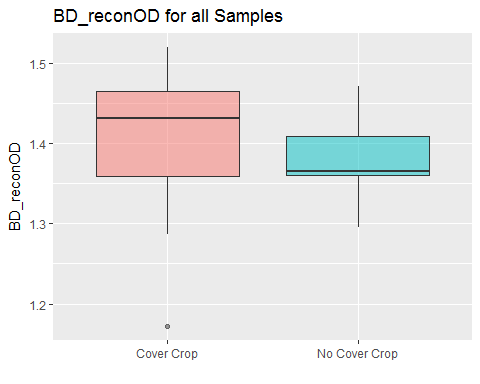
# [1] "WRC\_core13"



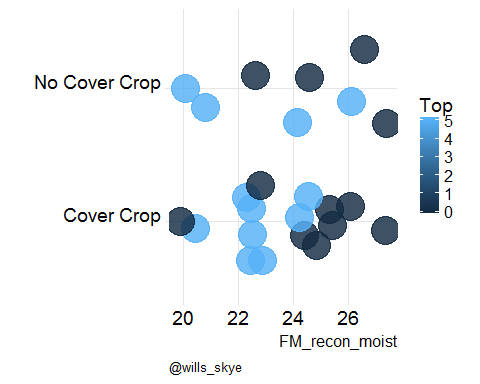
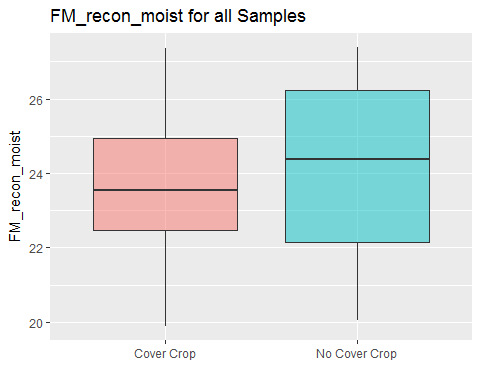
# [1] "BD\_recon13"



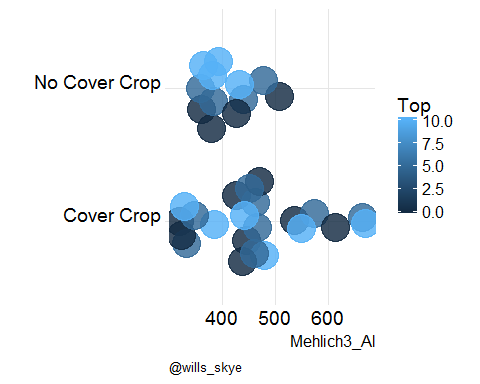
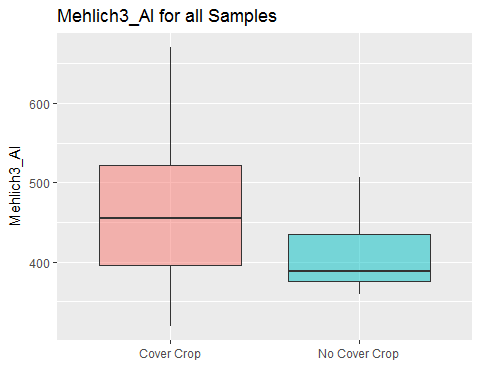
# [1] "BD\_reconOD"



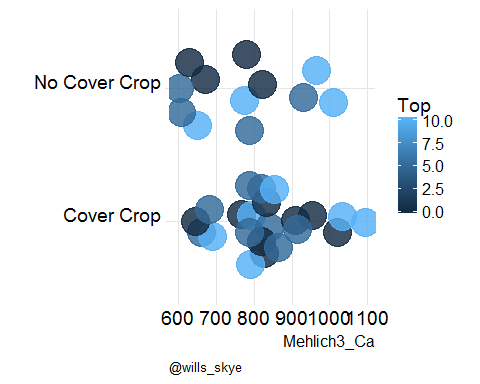
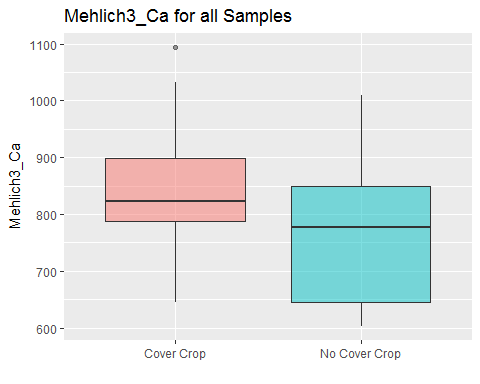
# [1] "FM\_recon\_moist"



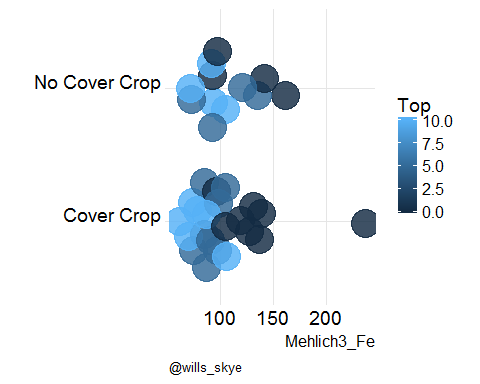
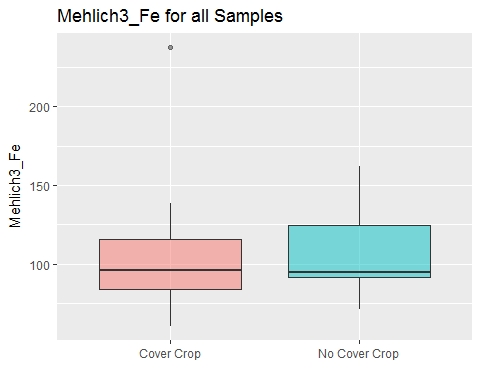
# [1] "Mehlich3\_Al"



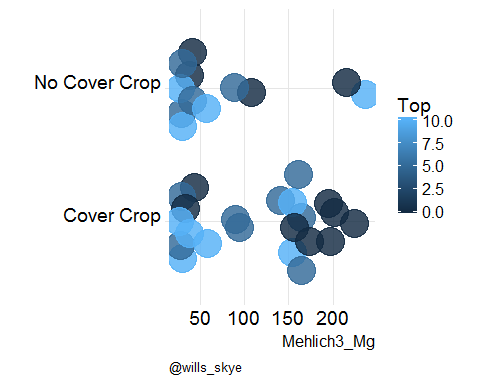
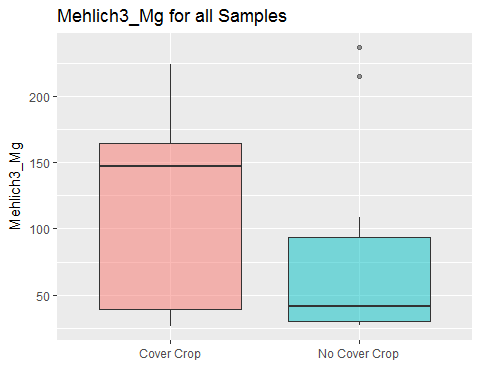
# [1] "Mehlich3\_Ca"



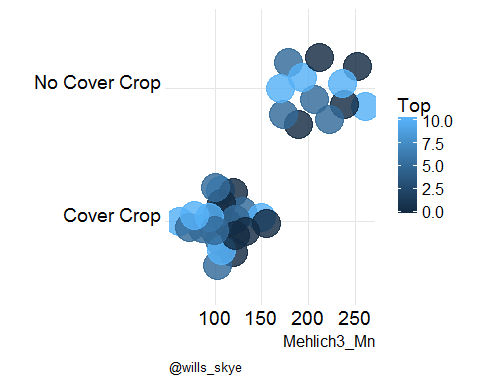
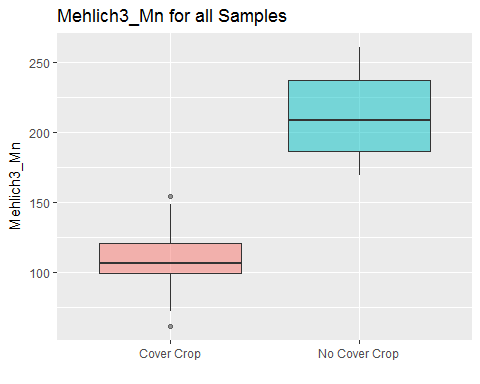
# [1] "Mehlich3\_Fe"



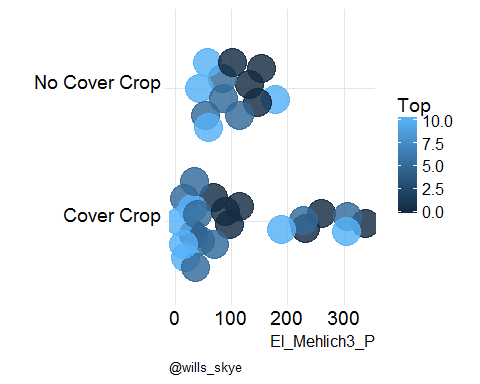
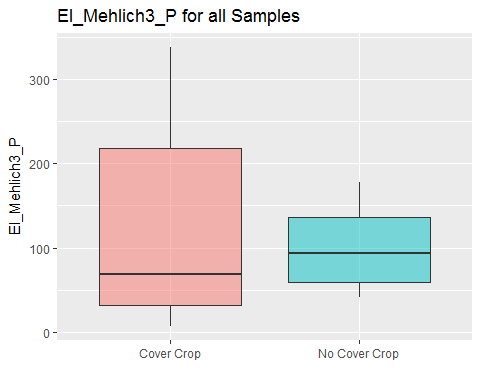
# [1] "Mehlich3\_Mg"



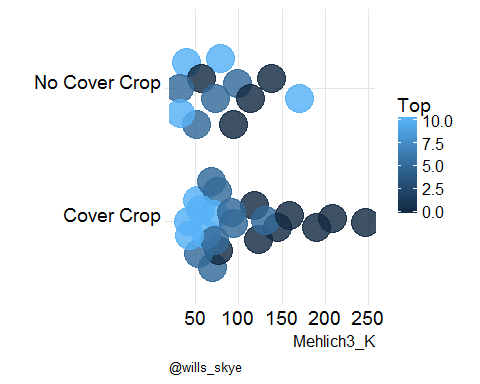
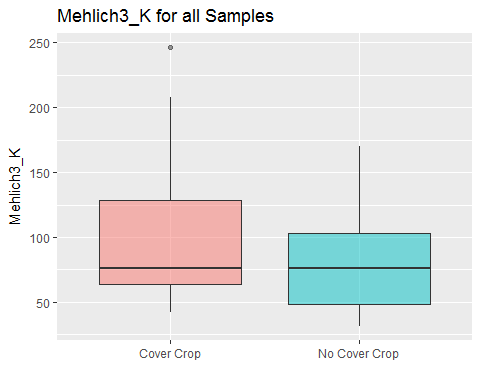
# [1] "Mehlich3\_Mn"



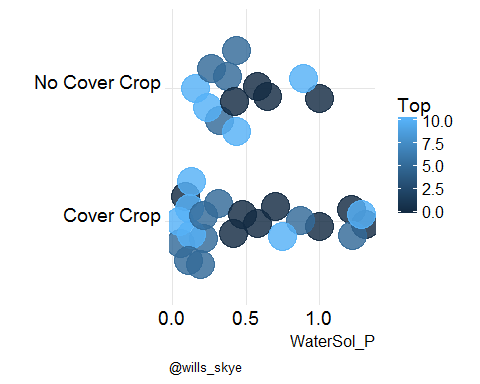
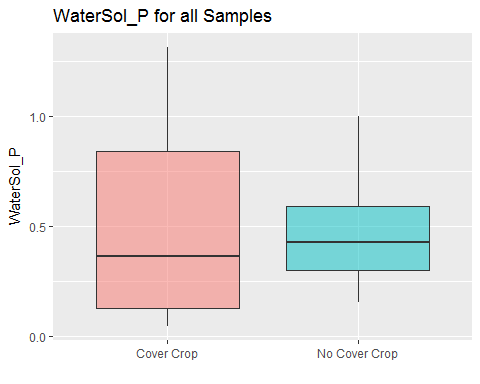
# [1] "El\_Mehlich3\_P"



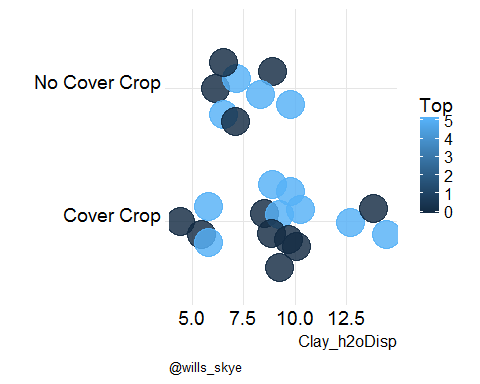
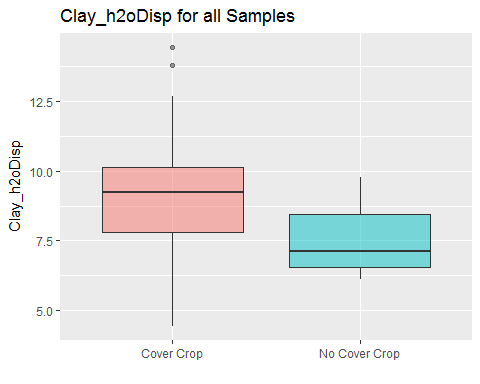
# [1] "Mehlich3\_K"



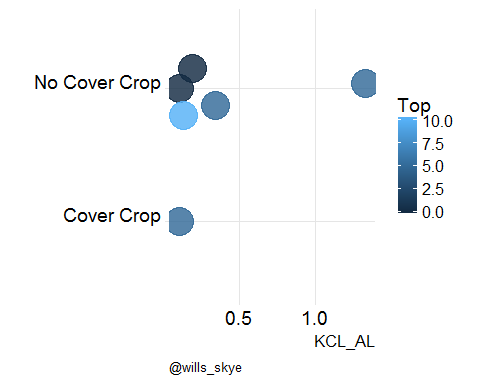
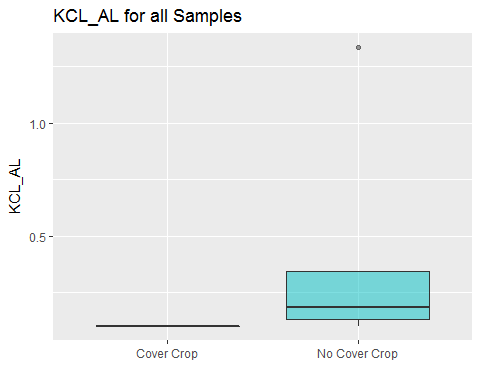
# [1] "WaterSol\_P"



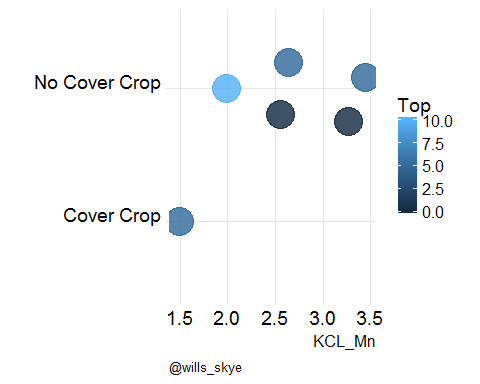
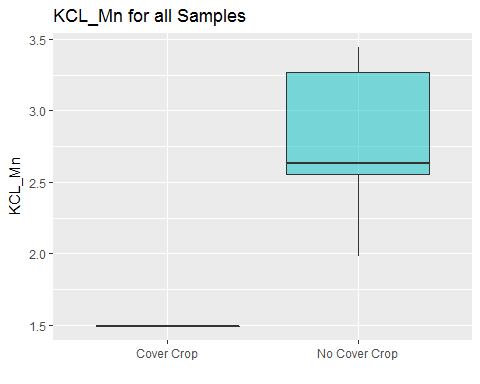
# [1] "Clay\_h2oDisp"



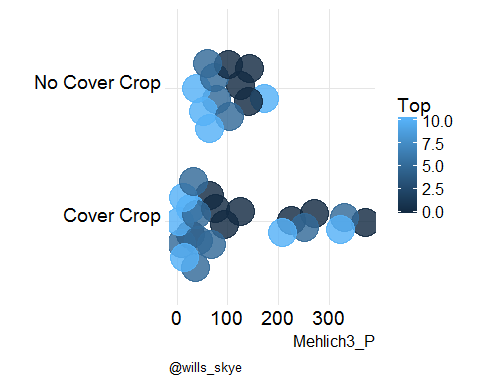
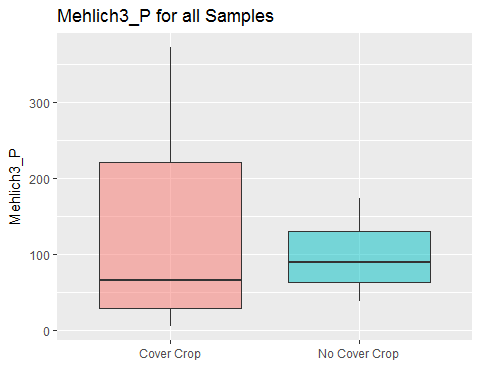
# [1] "KCL\_AL"



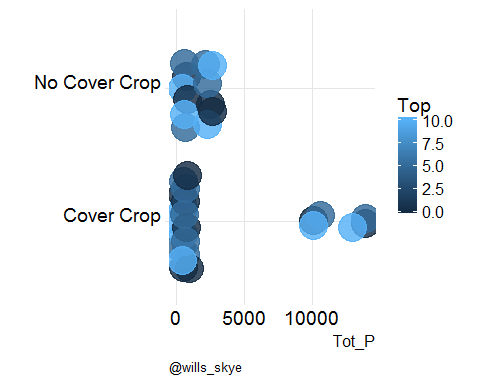
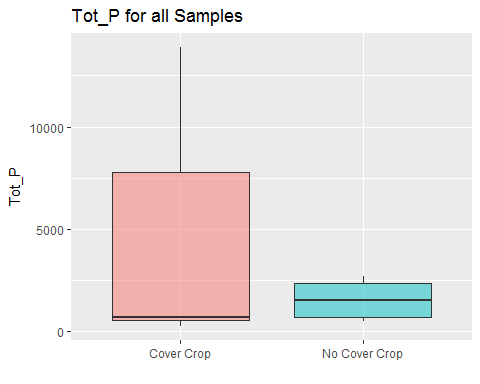
# [1] "KCL\_Mn"



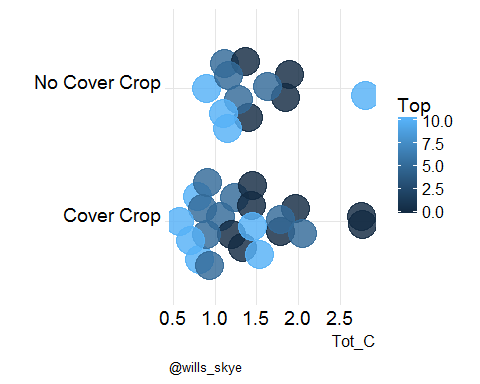
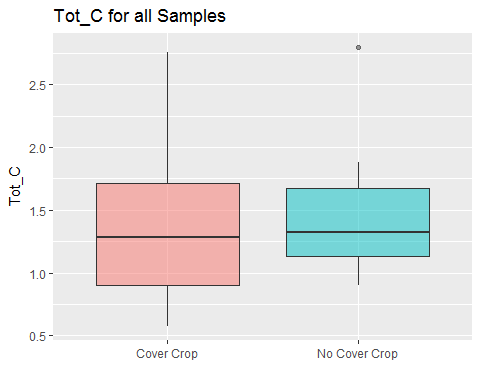
# [1] "Mehlich3\_P"



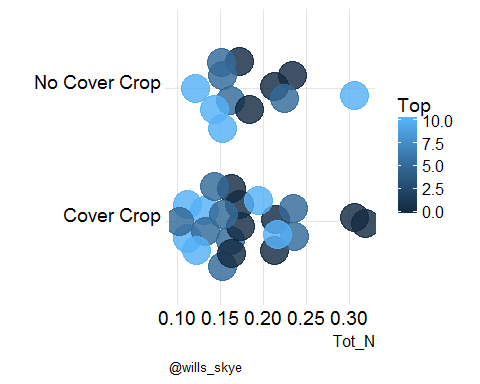
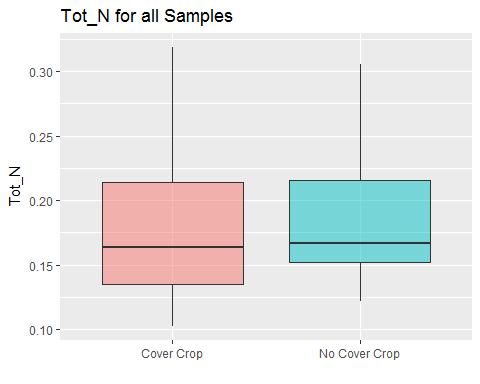
# [1] "Tot\_P"



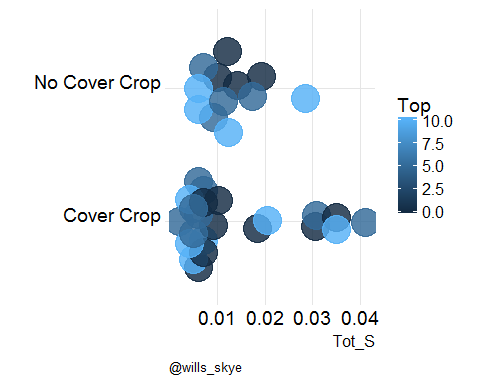
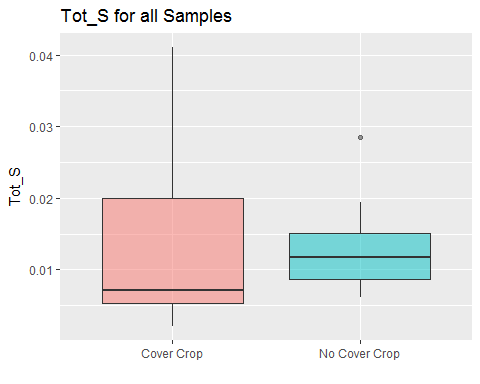
# [1] "Tot\_C"



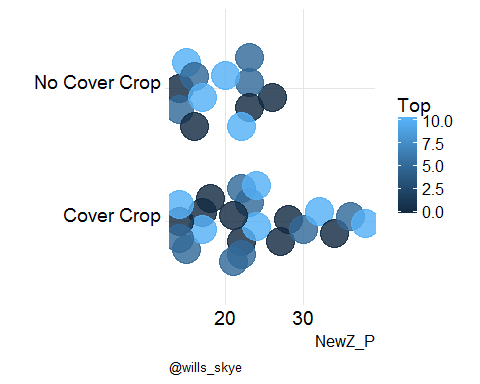
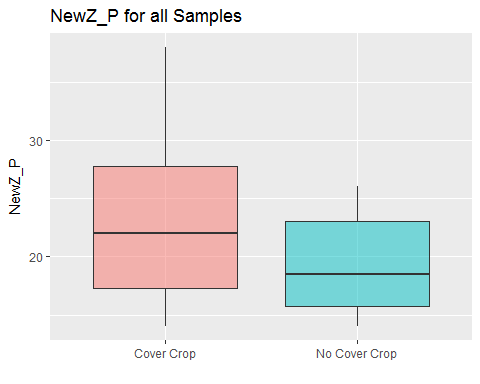
# [1] "Tot\_N"



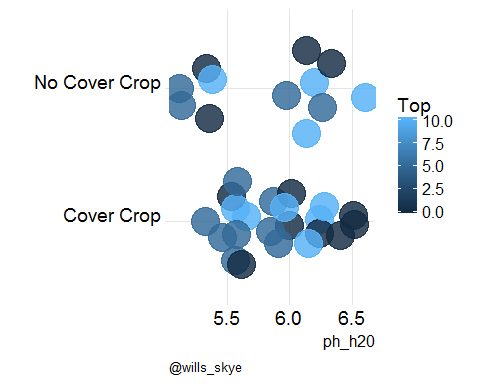
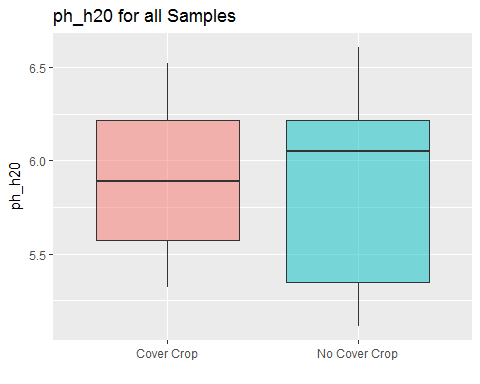
# [1] "Tot\_S"



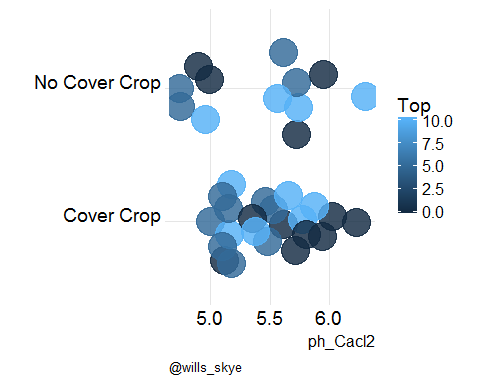
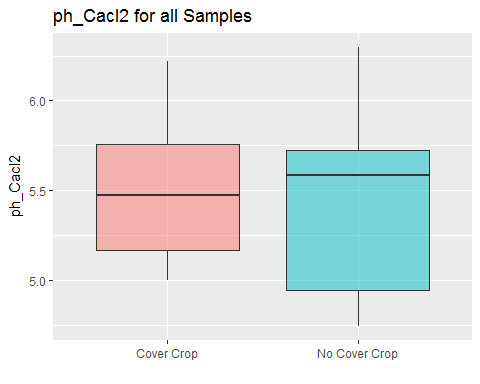
# [1] "NewZ\_P"



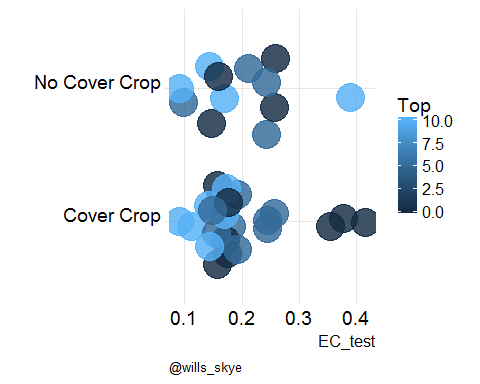
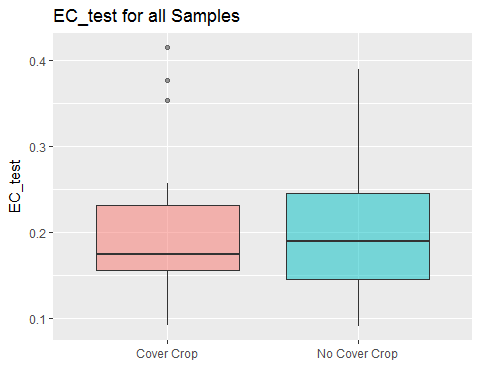
# [1] "ph\_h20"



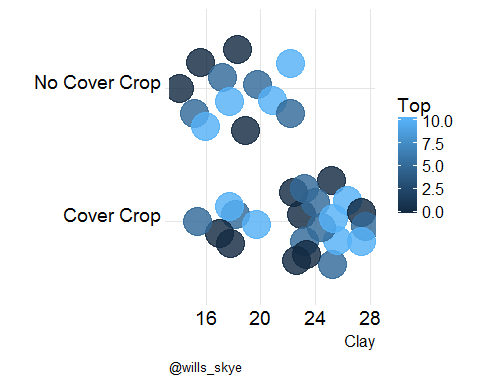
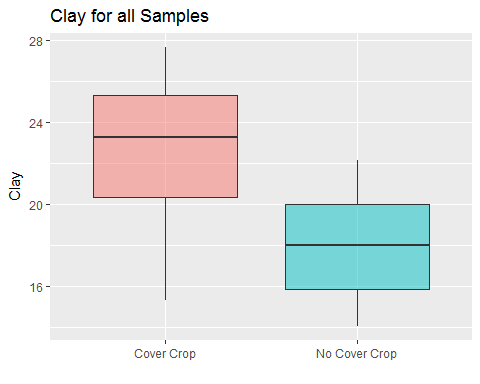
# [1] "ph\_Cacl2"



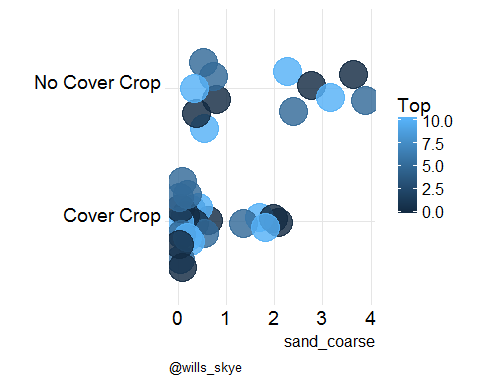
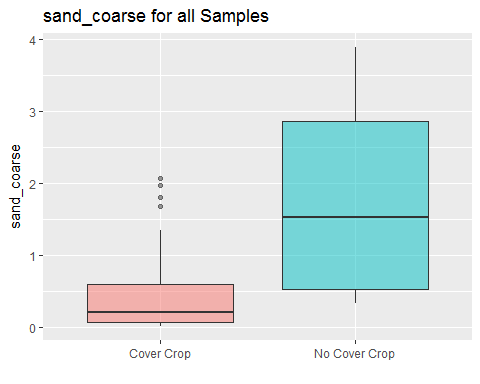
# [1] "EC\_test"



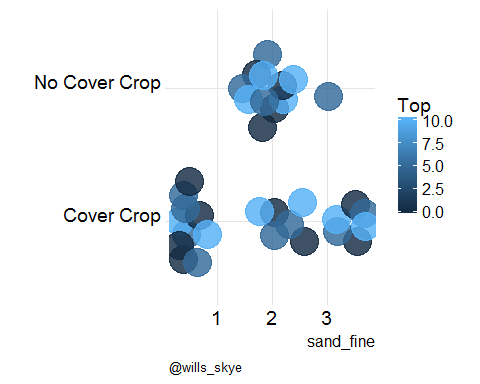
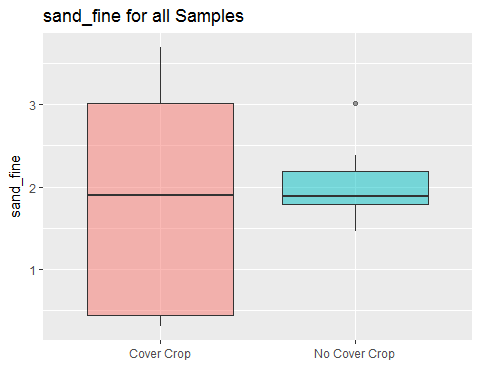
# [1] "Clay"



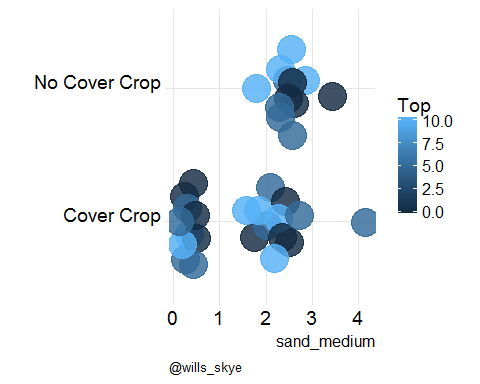
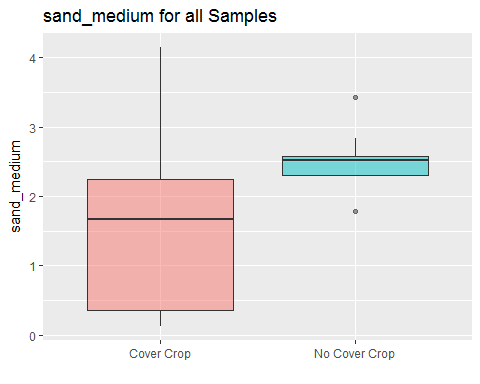
# [1] "sand\_coarse"



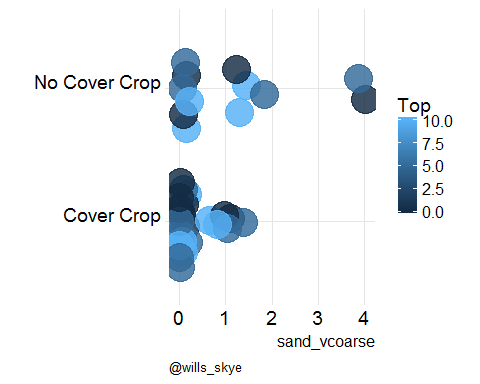
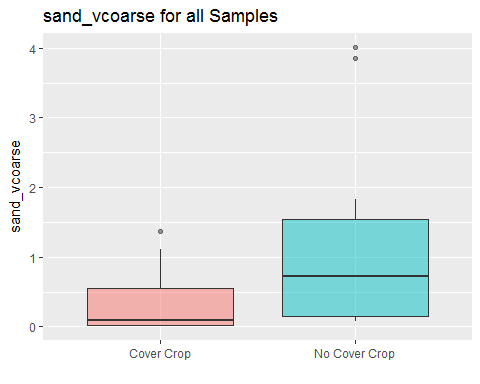
# [1] "sand\_fine"



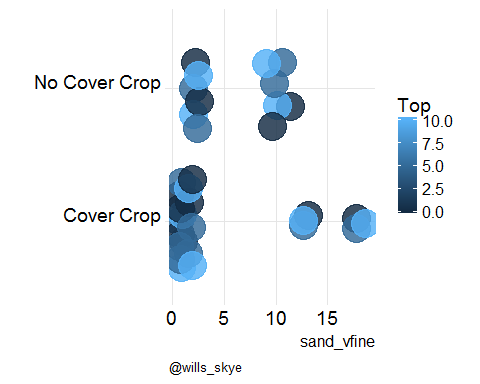
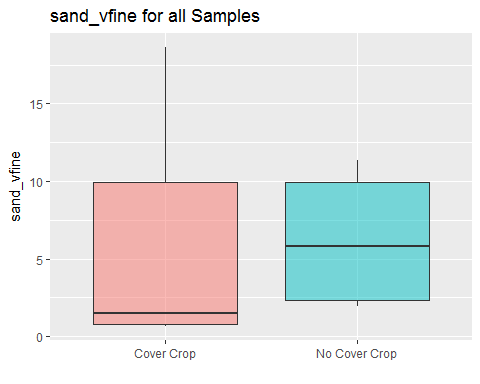
# [1] "sand\_medium"



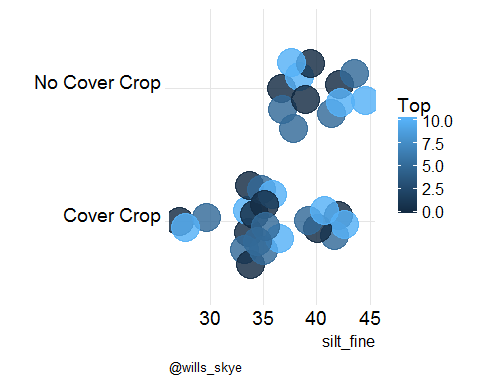
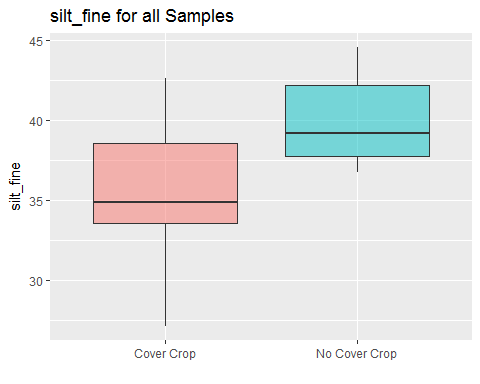
# [1] "sand\_vcoarse"



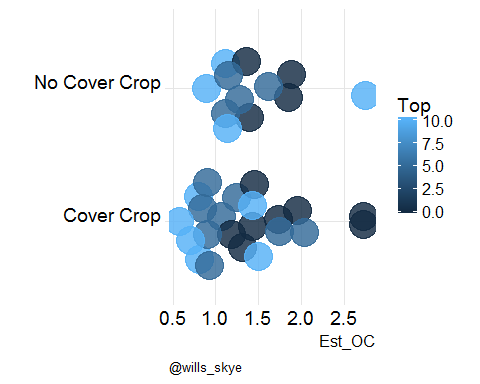
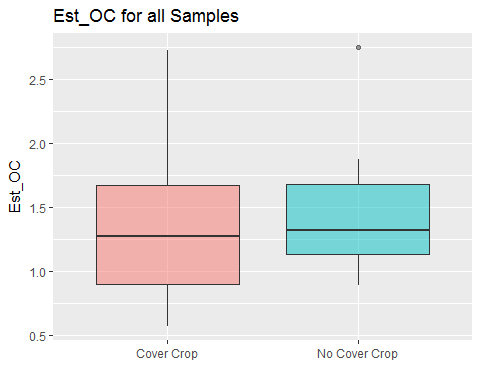
# [1] "sand\_vfine"



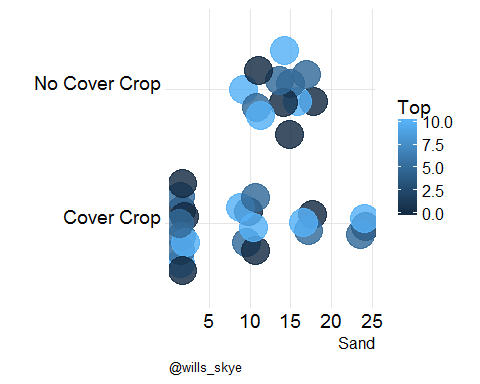
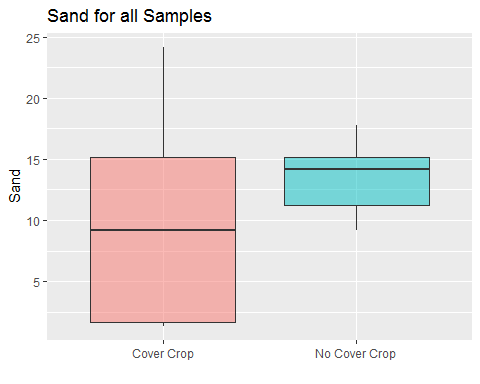
# [1] "silt\_fine"



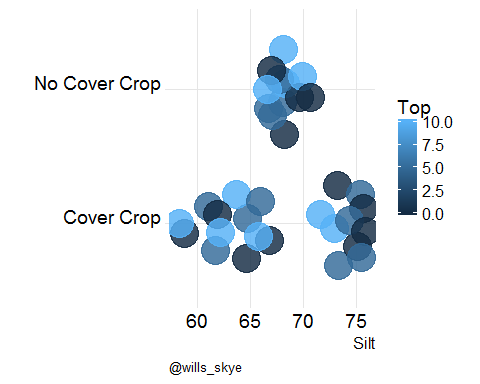
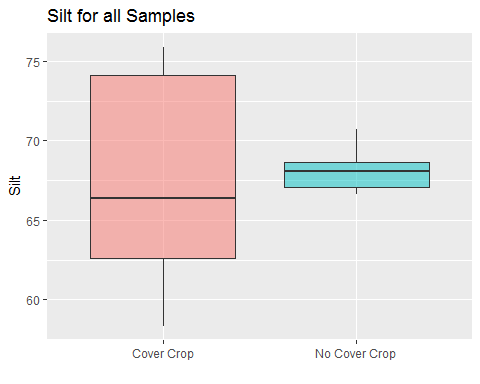
# [1] "Est\_OC"



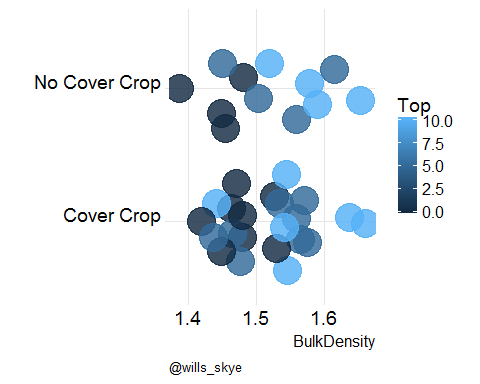
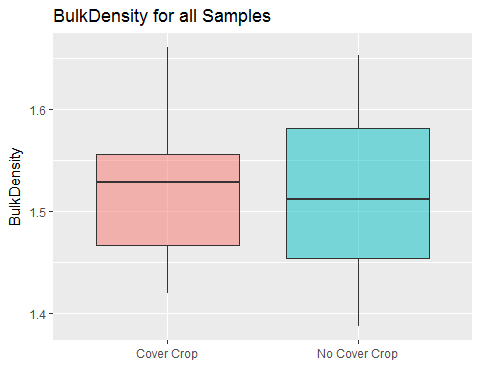
# [1] "Sand"



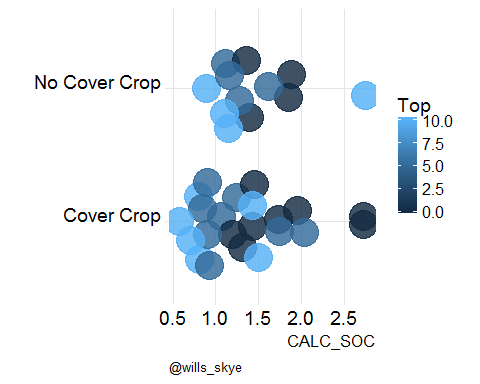
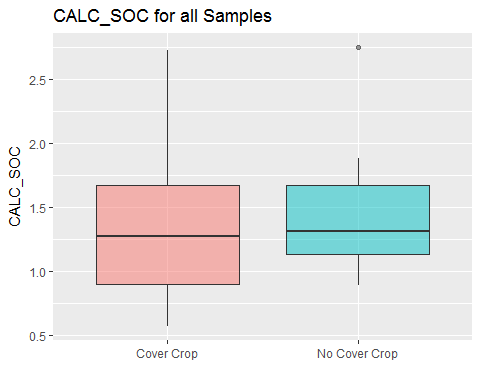
# [1] "Silt"



# [1] "BulkDensity"



# [1] "CALC\_SOC"



## Summary by comparable

# Test for diff

### by Surface, A horizons, and then B horizons

# Analysis of Variance Table  
#   
# Response: Value  
# Df Sum Sq Mean Sq F value Pr(>F)  
# Comp 1 57.04 57.042 0.6698 0.4322  
# Residuals 10 851.62 85.162

# # A tibble: 12 x 5  
# Prop data mod tidy n  
# <chr> <list> <list> <list> <int>  
# 1 AggStab <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 2 Bgluc <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 3 Pom\_C <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 4 Pom\_N <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 5 CEC\_ph7 <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 6 WaterSol\_P <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 7 Mehlich3\_P <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 8 Tot\_P <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 9 Tot\_N <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 10 NewZ\_P <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 11 BulkDensity <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12  
# 12 CALC\_SOC <tibble [12 x 14]> <anova [2 x 5]> <data.frame [2 x ~ 12

# # A tibble: 24 x 8  
# Prop n term df sumsq meansq statistic p.value  
# <chr> <int> <chr> <int> <dbl> <dbl> <dbl> <dbl>  
# 1 AggStab 12 Comp 1 57.0 5.70e+1 0.670 0.432   
# 2 AggStab 12 Residuals 10 852 8.52e+1 NA NA   
# 3 Bgluc 12 Comp 1 9165 9.17e+3 5.14 0.0469  
# 4 Bgluc 12 Residuals 10 17844 1.78e+3 NA NA   
# 5 Pom\_C 12 Comp 1 0.195 1.95e-1 9.63 0.0112  
# 6 Pom\_C 12 Residuals 10 0.203 2.03e-2 NA NA   
# 7 Pom\_N 12 Comp 1 0.000572 5.72e-4 3.71 0.0829  
# 8 Pom\_N 12 Residuals 10 0.00154 1.54e-4 NA NA   
# 9 CEC\_ph7 12 Comp 1 12.0 1.20e+1 1.39 0.265   
# 10 CEC\_ph7 12 Residuals 10 86.4 8.64e+0 NA NA   
# # ... with 14 more rows

## Alternate tests

### Surface Sample Tests

#### using mixed models