

# Cannabis Terroir Data Visualization - A representation of data collected by John Bershaw, PhD

Robert Bates, Melissa Palmer, Louis Tu, Luke Colten, Anika Walter

12/28/2021

## Statement of Problem

This project examines the data set provided by John Bershaw, PhD from a study that he is working on in order to see if local soils and growing conditions have an effect on the flavor profile or terroir of cannabis plants in the same fashion that they effect other similar plants like hops and grapes

*@bershaw*

## Methods

To aid in analysis of the data set, we first parsed out smaller subsets based on the unifying characteristics their variables held. This parsing was then leveraged by using these subsets as the denominations by which we generated graphs for all possible combinations of both within-group, and group-to-group, variables. By generating these plots we are able to quickly visually identify trends in data that may lead to further inquiry.

Spider plots were produced to better see the relationship between the data through the soil order and the soil subgroup classification.

ggplot and geom\_smooth were used to look for correlation between total terpenes and total THC as well as total terpenes and nitrogen in the soil.

Finally a correlation matrix was created using the Corrplot R package

*@Wei*

run on the entire data set for a visualization of all possible relationships, that may provide Dr. Bershaw with additional information. Several additional matrices were run on smaller subsets of data to tease out additional relationships.

## Results

Figure 1 below shows the relationship between ultisoil and terpene values

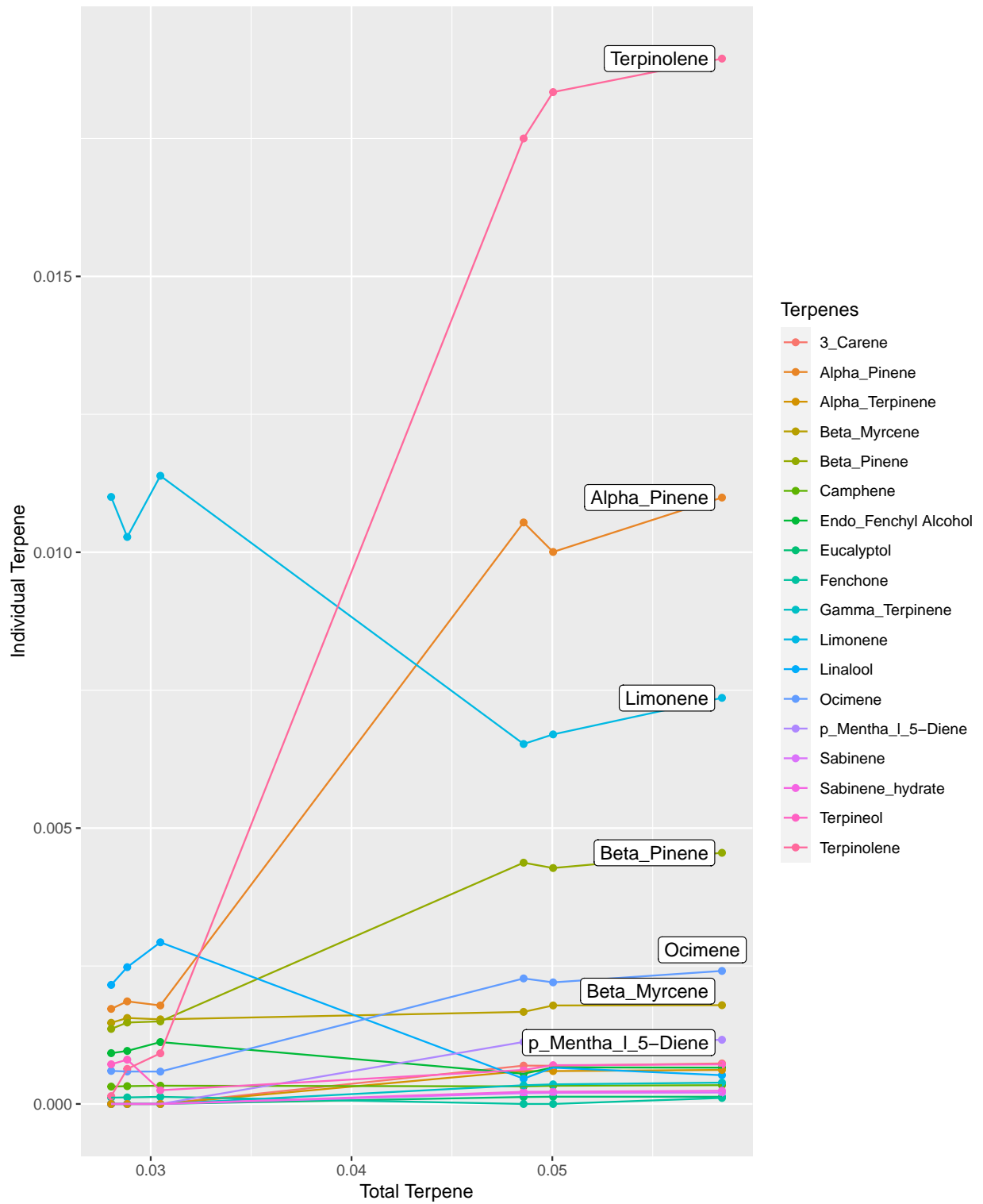


Figure 1. Total Terpene vs individual terpene values for Ultisols

Figure 2 below shows the relationship between mollisol and terpene values.

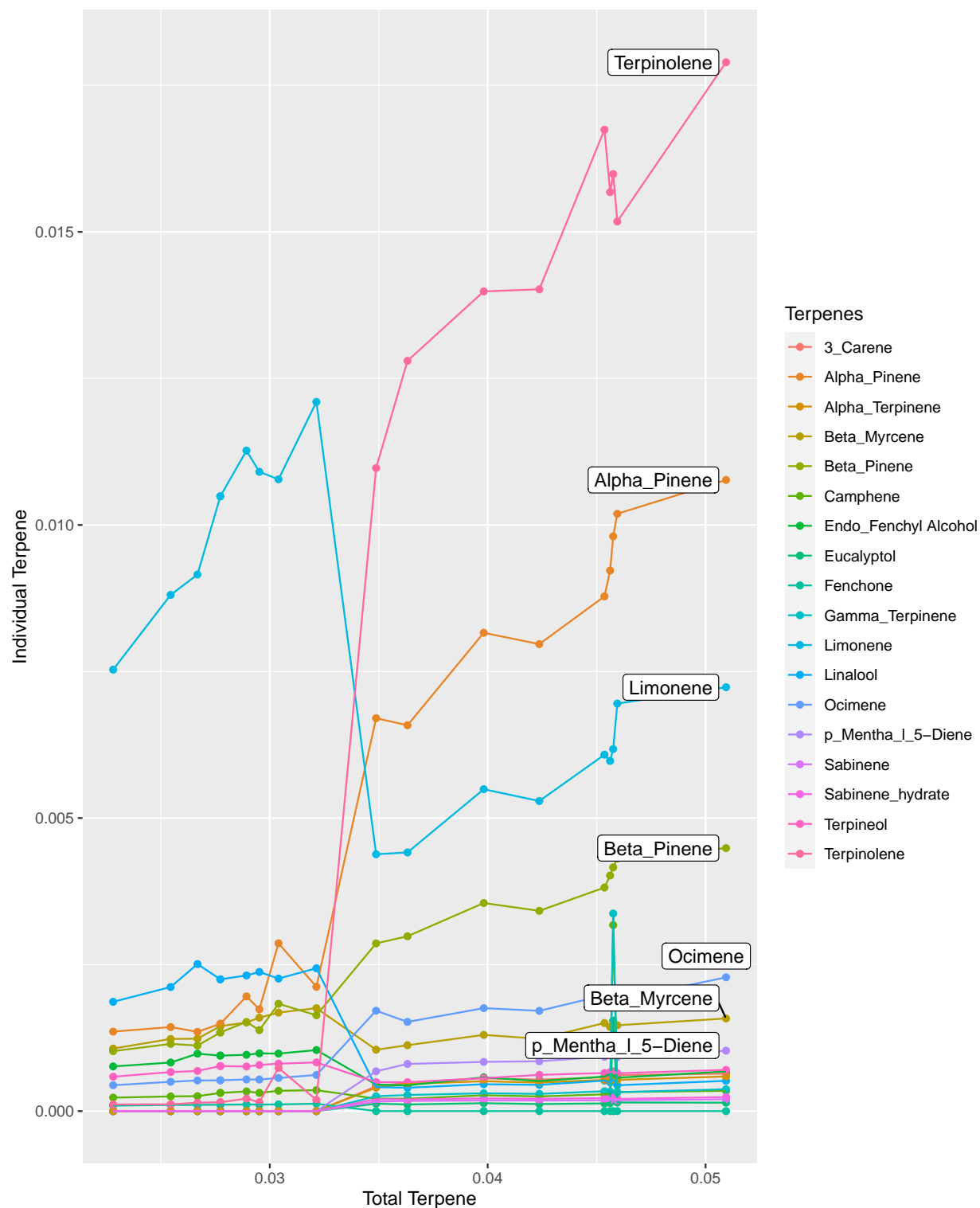


Figure 3 below shows the relationship between Alifosol and terpene values.

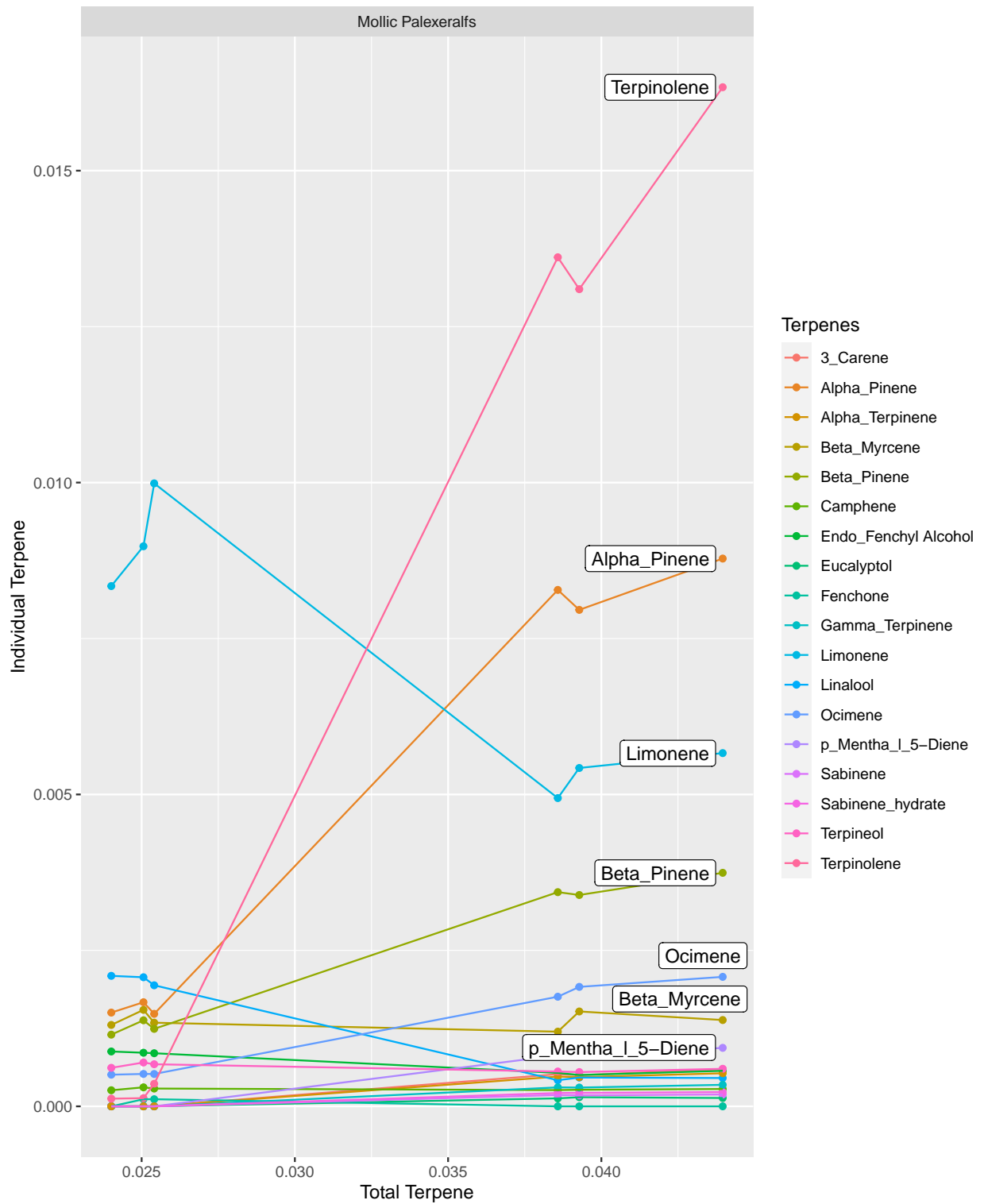
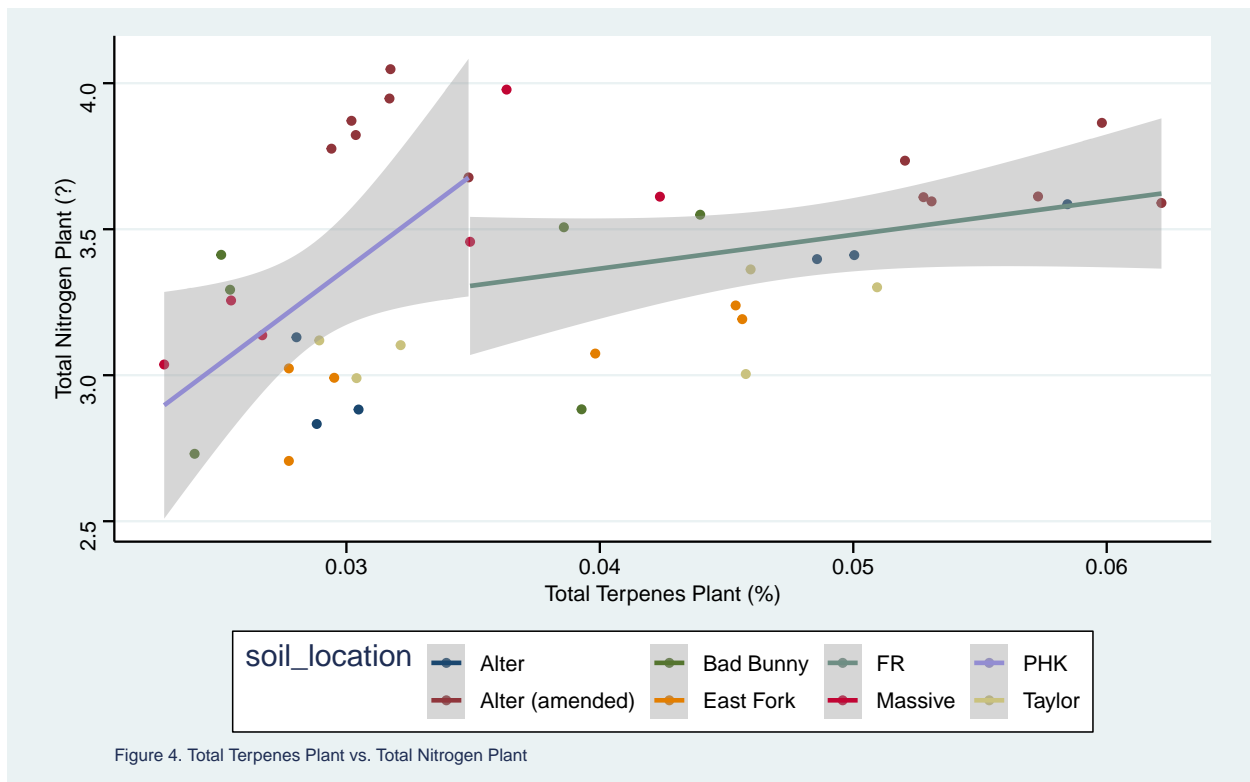


Figure 3. Total Terpene vs individual terpene values for Alfisols

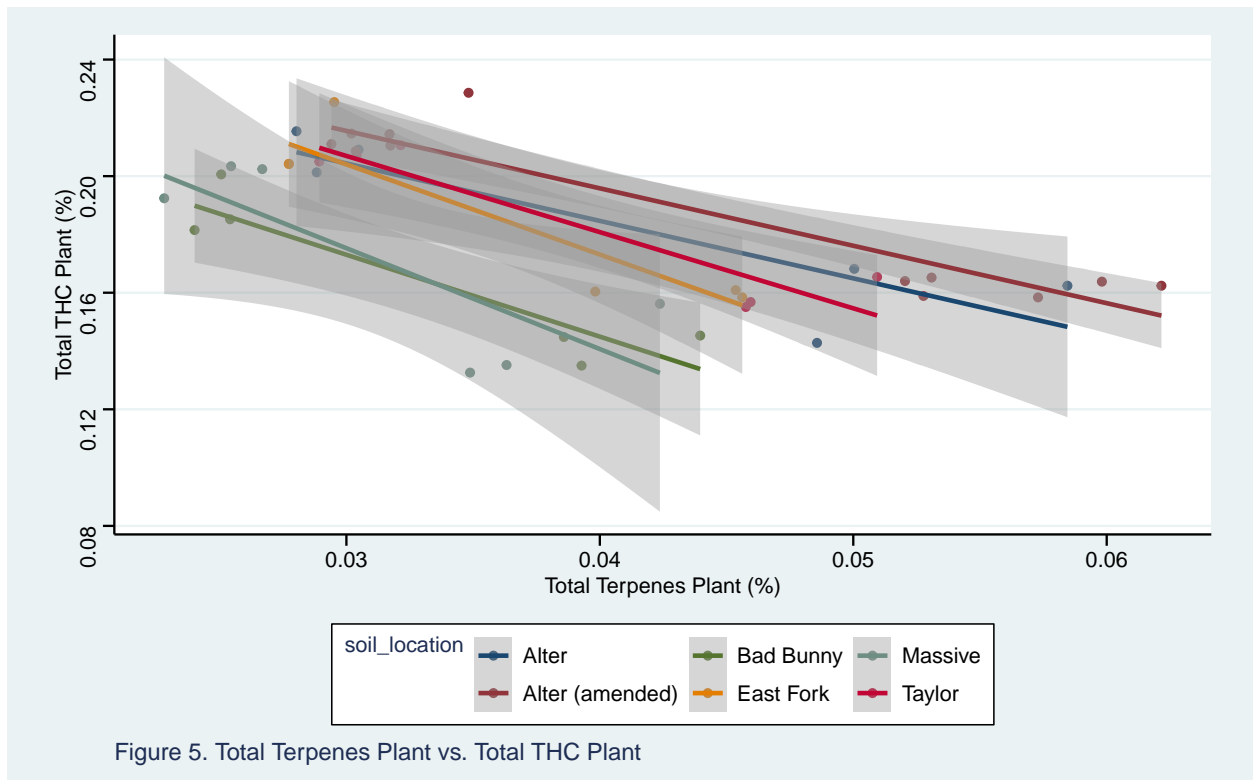
Figure 4 below shows the relationship between Nitrogen & Total Terpenes



Welch Two Sample t-test

data: `TableMastertotalterpenesplantandTableMastertotalnitrogenplant`  $t = -59.893$ ,  $df = 41.081$ ,  $p\text{-value} < 2.2e-16$  alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -3.439389 -3.215020 sample estimates: mean of x mean of y 0.03820262 3.36540714

Figure 5 below shows the relationship between total terpenes and total thc in the plants, plotted with geom\_smooth to show linear regression and standard deviation.



Welch Two Sample t-test

data: `TableMaster2totalterpenesplantandTableMaster2total_thc_plant`  $t = -30.292$ ,  $df = 53.758$ ,  $p\text{-value} < 2.2e-16$  alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval: -0.1518889 -0.1330297 sample estimates: mean of x mean of y 0.03820262 0.18066190

Figure 6 below shows a correlation matrix, showing all numerical values of the data plotted to try and find additional patterns within the data

## Warning in cor(IAMCurious4): the standard deviation is zero

Figure 6 – All data plotted in a correlation matrix

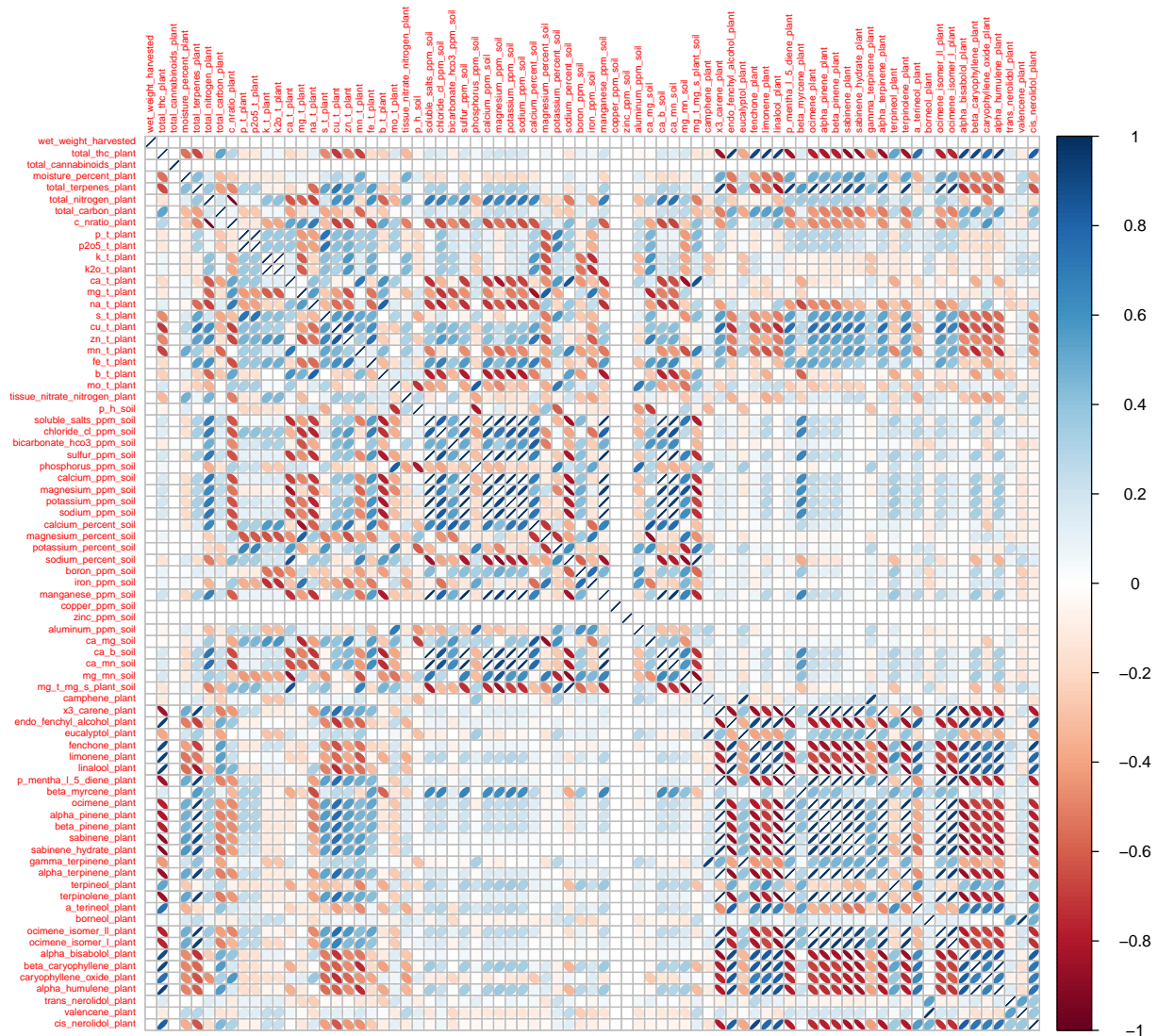


Figure 7 Shows monoterpenes vs. total carbon and nitrogen with a 95% CI.

**Figure 7. Monoterpenes vs Total Carbon & Nitrogen, 95% CI**

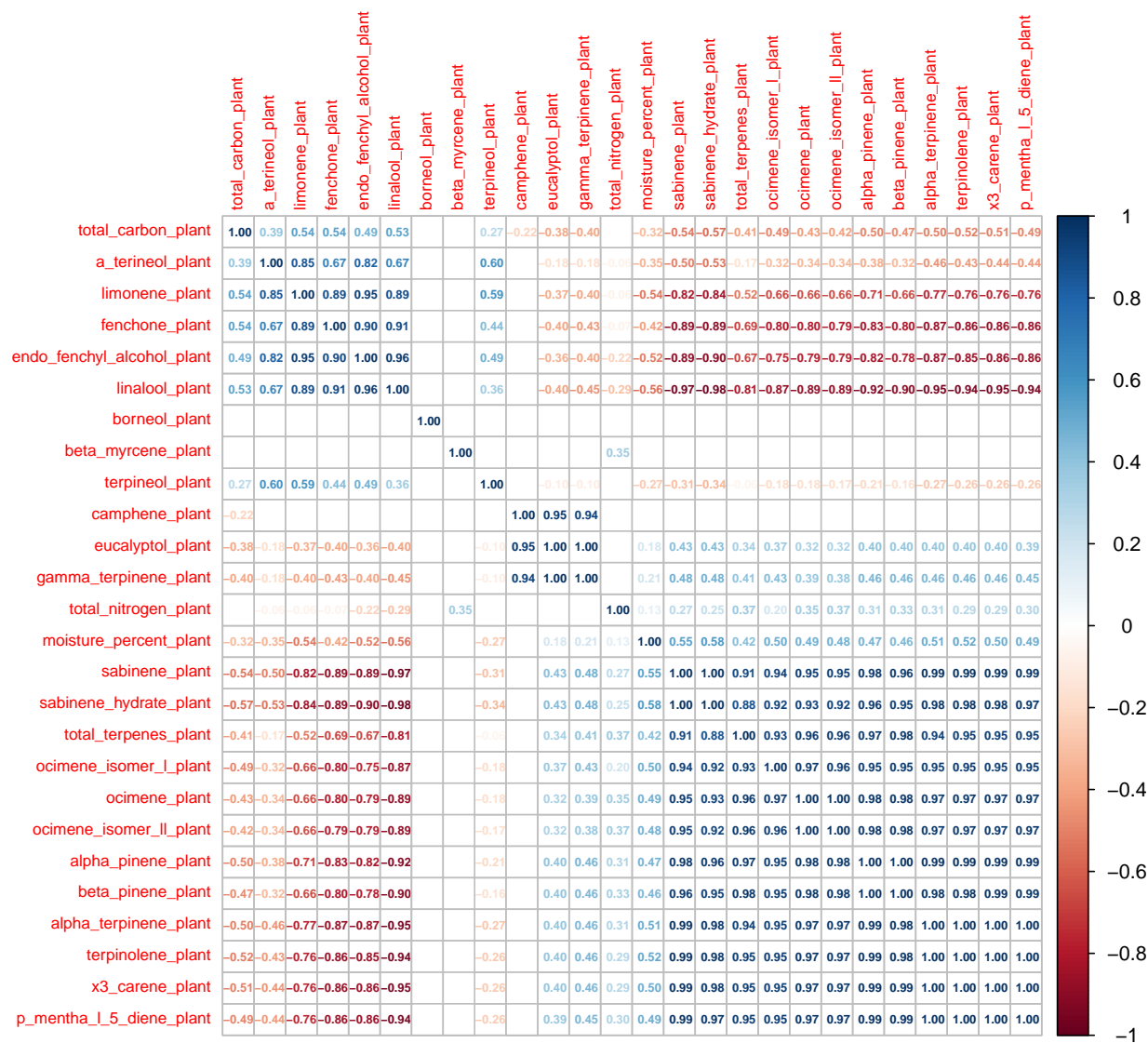




Figure 8 below shows soil nutrients against sesquiterpenes:

Figure 8. Soil Nutrients vs. Sesquiterpenes

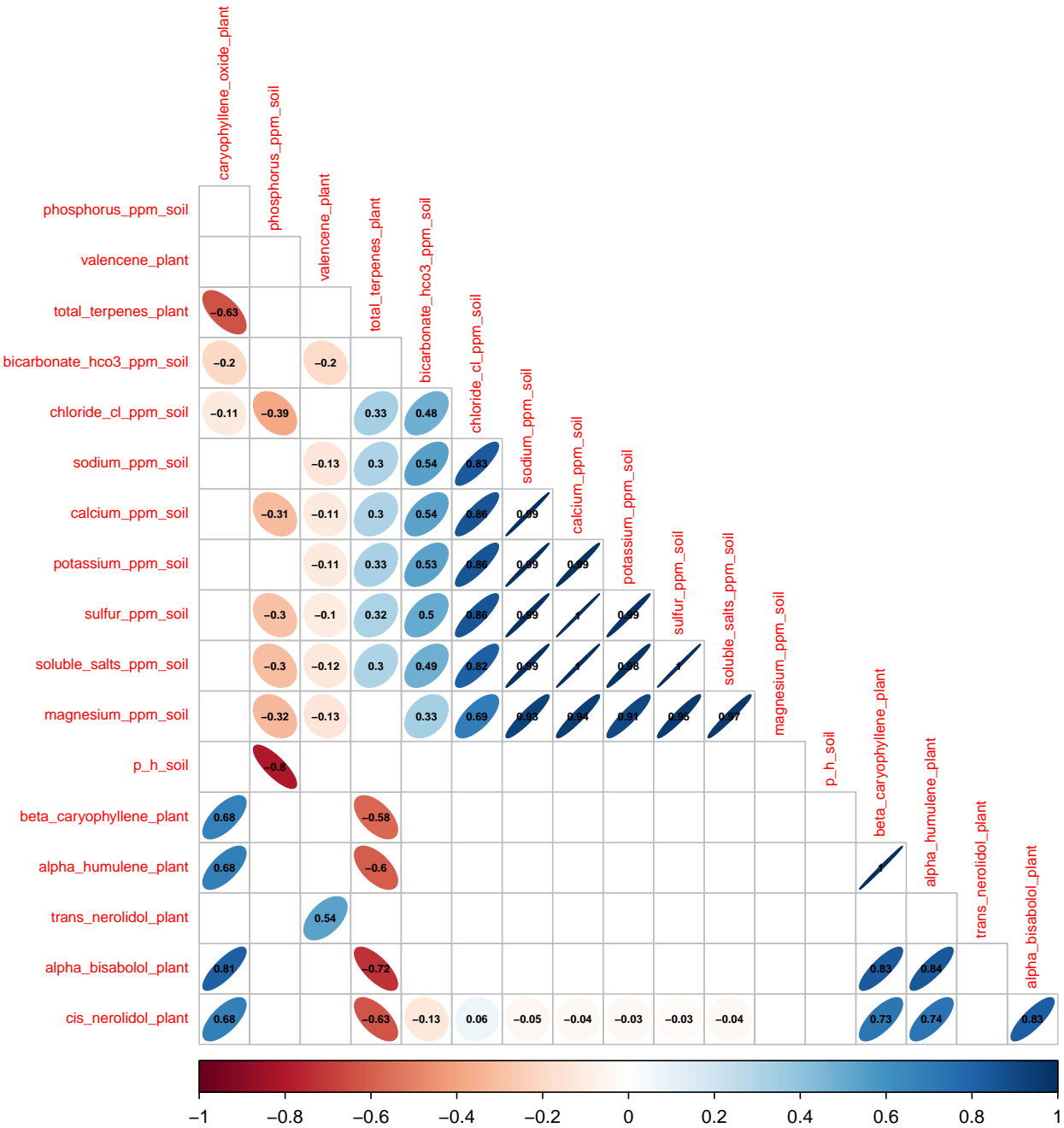


Figure 9 shows soil nutrients against monoterpenes:

### Figure 9. Soil Nutrients vs. Monoterpenes

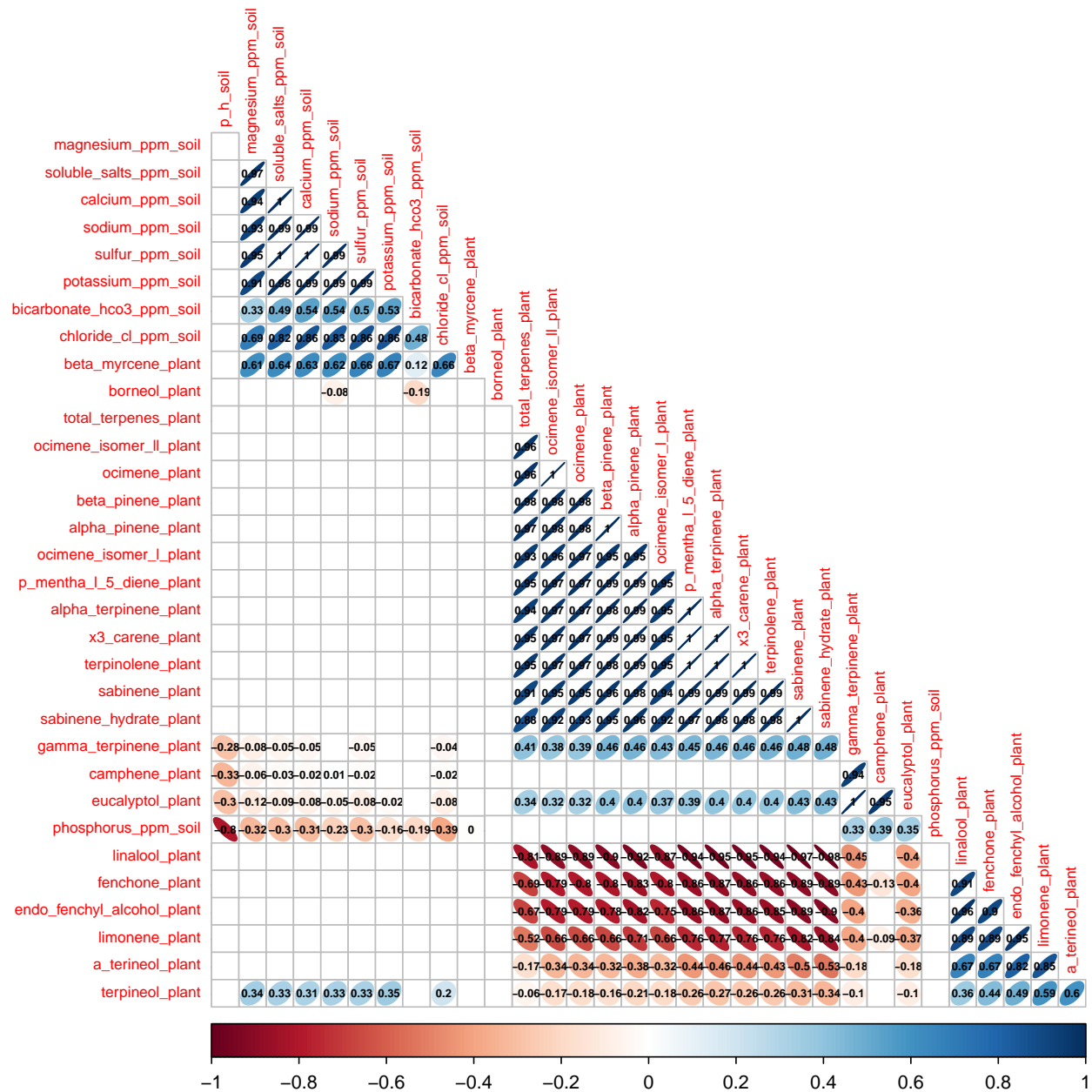


Figure 10 highlights strain FR against monoterpenes with a 95% CI:

Figure 10. FR vs. Monoterpenes

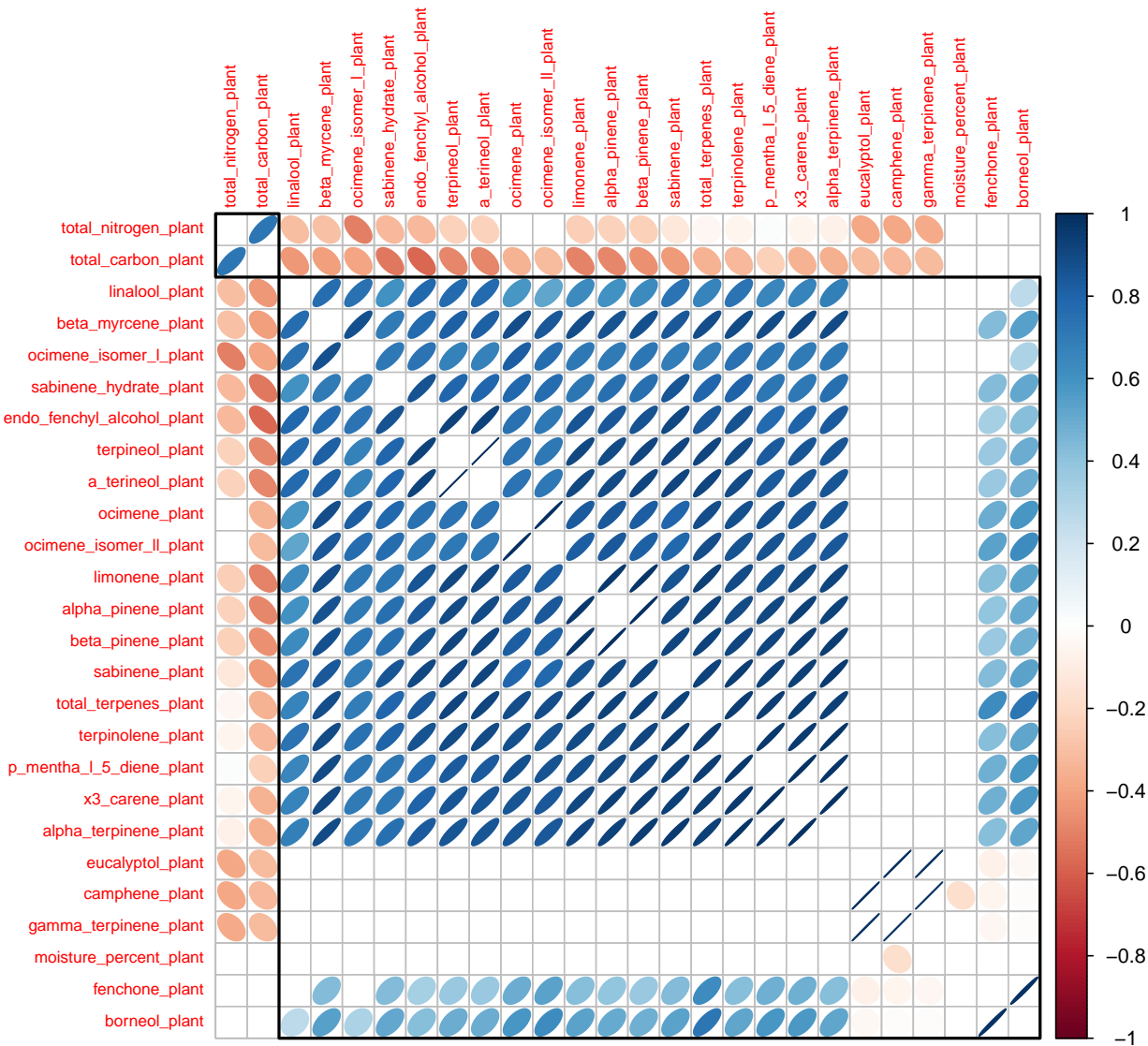


Figure 11 highlight strain FR against Sesquiterpenes with a 95% CI:

Figure 11. FR vs. Sesquiterpenes

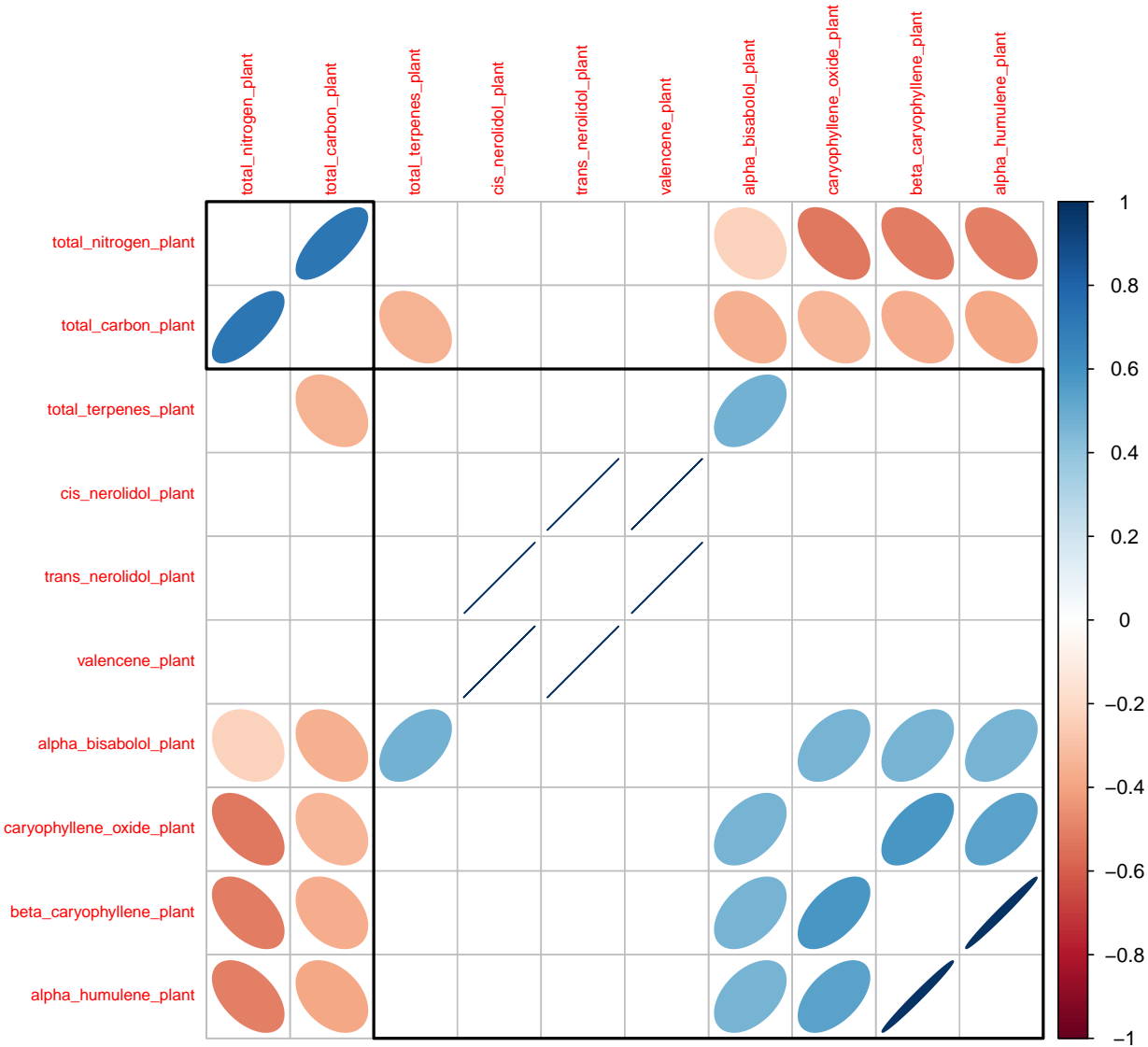


Figure 12 highlights the warning message from trying to make a correlation with tibble that have a standard deviation of zero:

```
## Warning in cor(IAMCurious2): the standard deviation is zero
## Warning in cor(IAMCurious4): the standard deviation is zero
## Warning in cor(PHKMonoTerps): the standard deviation is zero
## Warning in cor(SPHKTerps): the standard deviation is zero
## Warning in cor(MDirtTerpFROG): the standard deviation is zero
## Warning in cor(MDirtTerpPHK): the standard deviation is zero
## Warning in cor(SDirtTerpFROG): the standard deviation is zero
## Warning in cor(SDirtTerpPKG): the standard deviation is zero
```

## Discussion/Conclusion(s)

Patterns were seen in soil order and total terpenes versus individual terpene as shown in Figures 1, 2 and 3. There was a consistent and steep increase of terpinolene with total terpene values across Alfisols, Ultisols, and Mollisols. There was also an increase in Alpha pinene values with total terpene values across Alfisols, Ultisols, and Mollisols. An initially high value of limonene that generally increases then decreases with some variation across Alfisols, Ultisols, and Mollisols. Finally there were gradually increasing levels of beta pinene values with total terpenes across Alfisols, Ultisols, and Mollisols.

Total nitrogen of the soil when not separated by the strain was found to not be initially relevant on its effects on total terpene amounts (Figure 4). Due to this, Total nitrogen of the soil was not tested against individual terpenes. Total nitrogen of the soil was also found not to have a correlation with total THC amounts (Figure 5). This was seen by the lack of a pattern within the plotted graphs. The loop created to compare two different tibbles, showed unexpectedly very little evidence of any data showing an important trend between any of the variables discussed above.

The correlation matrices were created to see if other relationships between all variables included in our joined master tibble. Figure 6 shows all numerical data within the dataset of the correlation matrix. Additional tibbles were created to narrow down the results by strain. Figures 8 and 9 highlight the correlations between soil data against sesquiterpenes and monoterpenes respectively. Interestingly, the correlation matrices set up to look at the data by strain then PHK, when plotted by soil data and monoterpenes had a standard deviation of zero, and when plotted by soil data and sesquiterpenes also had a standard deviation of zero, however FR showed some correlation and is plotted in in Figures 10 and 11.

## References

Bershaw, John, Kyle White, Bethany Carder, Scott Burns, Cody Alter, Jason Rambo, Eric Wendt, et al. "Cannabis Terroir: The Influence of Soil on Plant Chemistry for Cannabis Sativa Grown Outdoors in Southern Oregon." Manuscript. Oregon, not published. Wei, Taiyun, and William Simko. Corplot Introduction. Accessed March 14, 2022. <https://cran.r-project.org/web/packages/corplot/vignettes/corplot-intro.html>.

## Appendix 1 - R code

```
##-----Setup Code-----

#import libs
library(tidyverse)
library(google sheets4)
library(RColorBrewer)
library(janitor)
library(corrplot)
library(Hmisc)
library(kableExtra)
library(ggcorrplot)
library(sm)
library(readxl)
library(ggrepel)
library(ggthemes)

#this code is my file save path which I have given the name of 'savefile'.
savefile = "G:/My Drive/G324/Assignment 5 code"

#theme data
PlotColorScheme <- hcl.colors(256, palette = "Plasma")

#this code is to login to my google account. Remember to type the 1 down
#in the console before running anything else
gs4_auth('mepalmer@pdx.edu')

##-----Functions Code-----
#Credit to Bobby for writing the functions to work with our tibbles
#plot all combinations of tibble columns
testOneTibble <- function(tibbleOne) {
  tibVarList <- colnames(tibbleOne)
  tibLength <- length(tibbleOne)
  for (i in 1:(tibLength - 1)) {
    for (j in (i + 1):tibLength) {
      subText <- paste(tibVarList[i], "vs", tibVarList[j])
      plot (
        tibbleOne %>%
          ggplot(aes(
            x = get(tibVarList[i]), y = get(tibVarList[j])
          )) +
          geom_point(show.legend = FALSE) +
          scale_color_gradientn(colors = PlotColorScheme) +
          labs(
            x = tibVarList[i],
            y = tibVarList[j],
            subtitle = subText
          )
      )
    }
  }
}
```

```

#plot all combinations of columns from two separate tibbles
testTwoTibbles <- function(tibbleOne, tibbleTwo) {
  varListOne <- colnames(tibbleOne)
  varListTwo <- colnames(tibbleTwo)
  lengthOne <- length(tibbleOne)
  lengthTwo <- length(tibbleTwo)
  heightOne <- nrow(tibbleOne)
  heightTwo <- nrow(tibbleTwo)
  joinedTibble <-
    left_join(tibble(tibbleOne, key = 1:heightOne),
              tibble(tibbleTwo, key = 1:heightTwo))
  view(joinedTibble)
  for (i in 1:lengthOne) {
    for (j in 1:lengthTwo) {
      subText <- paste(varListOne[i], "vs", varListTwo[j])
      plot (
        joinedTibble %>%
          ggplot(aes(
            x = get(varListOne[i]), y = get(varListTwo[j])
          )) +
          geom_point(show.legend = FALSE) +
          scale_color_gradientn(colors = PlotColorScheme) +
          labs(
            x = varListOne[i],
            y = varListTwo[j],
            subtitle = subText
          )
      )
    }
  }
}

##-----end Functions-----

##-----Data1-----

#This is data call number one, Louis and Bobby each queried the dataset differently

#this code is to establish that the object Puffmaster is my url for my google
#sheet of interest,
#"https://docs.google.com/spreadsheets/d/1sup9_iIEcPW5ulxAtzm9BIDD8LWTANslN0sUKrM7l5s"

puffmaster <- "https://docs.google.com/spreadsheets/d/1sup9_iIEcPW5ulxAtzm9BIDD8LWTANslN0s
smoke <- read_sheet(puffmaster,
  range = 'Master Sheet (corrected)!A3:CD45', .name_repair=unique)
#view(smoke)

dirt <- read_sheet(puffmaster, range= ('Soil_Info!A1:F6'), .name_repair = unique)
#view(dirt)

SoilData <- smoke %>%
  #select(Soil, PlantID, PlanterPosition, "pH (soil)": "Mg_T/Mg_s (plant/soil)") %>%
  rename("WetWeight" = "WetWeight(harvested)",
        "Total_THT" = "TotalTHC(plant)",

```

"Total\_Cannabinoids" = "TotalCannabinoids(plant)",  
 "Moisture\_Percent" = "Moisture%(plant)",  
 "Total\_Terpenes" = "TotalTerpenes(plant)",  
 "Total\_Nitrogen" = "TotalNitrogen(plant)",  
 "Total\_Carbon" = "Total Carbon (plant)",  
 "C/NRatio" = "C/NRatio (plant)",  
 "P\_T" = "P\_T (plant)",  
 "P205\_T" = "P205\_T (plant)",  
 "K\_T" = "K\_T(plant)",  
 "K20\_T" = "K20\_T (plant)",  
 "Ca\_T" = "Ca\_T (plant)",  
 "Mg\_T" = "Mg\_T (plant)",  
 "Na\_T" = "Na\_T (plant)",  
 "S\_T" = "S\_T (plant)",  
 "Cu\_T" = "Cu\_T (plant)",  
 "Zn\_T" = "Zn\_T (plant)",  
 "Mn\_T" = "Mn\_T (plant)",  
 "Fe\_T" = "Fe\_T (plant)",  
 "B\_T" = "B\_T (plant)",  
 "Mo\_T" = "Mo\_T (plant)",  
 "Tissue\_Nitrate\_Nitrogen" = "Tissue Nitrate Nitrogen (plant)",  
 "pH" = "pH (soil)",  
 "Soluble\_Salts" = "Soluble Salts (ppm) (soil)",  
 "Chloride(Cl)" = "Chloride (Cl) (ppm) (soil)",  
 "Bicarbonate(HCO3)" = "Bicarbonate (HCO3) (ppm) (soil)",  
 "Sulfur" = "Sulfur (ppm) (soil)",  
 "Phosphorus" = "Phosphorus (ppm) (soil)",  
 "Calcium" = "Calcium (ppm) (soil)",  
 "Magnesium" = "Magnesium (ppm) (soil)",  
 "Potassium" = "Potassium (ppm) (soil)",  
 "Sodium" = "Sodium (ppm) (soil)",  
 "Calcium\_Percent" = "Calcium (%) (soil)",  
 "Magnesium\_Percent" = "Magnesium (%) (soil)",  
 "Potassium\_Percent" = "Potassium (%) (soil)",  
 "Sodium\_Percent" = "Sodium (%) (soil)",  
 "Boron" = "Boron (ppm) (soil)",  
 "Iron" = "Iron (ppm) (soil)",  
 "Manganese" = "Manganese (ppm) (soil)",  
 "Copper" = "Copper (ppm) (soil)",  
 "Zinc" = "Zinc (ppm) (soil)",  
 "Aluminum" = "Aluminum (ppm) (soil)",  
 "Ca/Mg" = "Ca/Mg (soil)",  
 "Ca/B" = "Ca/B (soil)",  
 "Ca/Mn" = "Ca/Mn (soil)",  
 "Mg/Mn" = "Mg/Mn (soil)",  
 "Mg\_T/Mg\_s" = "Mg\_T/Mg\_s (plant/soil)",  
 "Camphene" = "Camphene (plant)",  
 "3\_Carene" = "3-Carene (plant)",  
 "Endo\_Fenchyl Alcohol" = "Endo-Fenchyl Alcohol (plant)",  
 "Eucalyptol" = "Eucalyptol (plant)",  
 "Fenchone" = "Fenchone (plant)",  
 "Limonene" = "Limonene (plant)",  
 "Linalool" = "Linalool (plant)",  
 "p\_Mentha\_l\_5-Diene" = "p-Mentha-l, 5-Diene (plant)",



```

"Beta_Myrcene" = "Beta-Myrcene (plant)",
"Ocimene" = "Ocimene (plant)",
"Alpha_Pinene" = "Alpha-Pinene (plant)",
"Beta_Pinene" = "Beta-Pinene (plant)",
"Sabinene" = "Sabinene (plant)",
"Sabinene_hydrate" = "Sabinene hydrate (plant)",
"Gamma_Terpinene" = "Gamma-Terpinene (plant)",
"Alpha_Terpinene" = "Alpha-Terpinene (plant)",
"Terpineol" = "Terpineol (plant)",
"Terpinolene" = "Terpinolene (plant)",
"A_Terineol" = "A-Terineol (plant)",
"Borneol" = "Borneol (plant)",
"Ocimene_isomer_ll" = "Ocimene isomer ll (plant)",
"Ocimene_isomer_l" = "Ocimene isomer l (plant)",
"Alpha-Bisabolol" = "Alpha-Bisabolol (plant)",
"Beta_Caryophyllene" = "Beta-Caryophyllene (plant)",
"Caryophyllene_Oxide" = "Caryophyllene Oxide (plant)",
"Alpha_Humulene" = "Alpha-Humulene (plant)",
"Trans_Nerolidol" = "Trans-Nerolidol (plant)",
"Valencene" = "Valencene (plant)",
"Cis_Nerolidol" = "Cis-Nerolidol (plant)")
# view(SoilData)

PlanterDiagram <- read_sheet(puffmaster, range = "Planter Diagram!B20:C28",
  .name_repair = "unique")
PlanterDiagram <- PlanterDiagram[-c(6),] %>%
  rename("Farm" = "...2")

#Join PlanterDiagram table and dirt table
SoilInfo <- full_join(PlanterDiagram, dirt)
#view(SoilInfo)

# FullTable <- left_join(smoke, SoilInfo, by = c("Soil" = "Raised Beds")) %>%
#   relocate(c("Farm":"Soil Series"), .after = PlanterPosition)
FullTable <- left_join(SoilData, SoilInfo, by = c("Soil" = "Raised Beds")) %>%
  relocate(c("Farm":"Soil Series"), .after = PlanterPosition) %>%
  rename("Soil_Order" = "Soil Order",
    "Soil_Subgroup" = "Soil Subgroup",
    "Soil_Series" = "Soil Series")
#Terpenes
AllTerpenesTable <- FullTable %>%
  select("Soil":"Farm",
    "Soil_Order",
    "Soil_Subgroup",
    "Total_Terpenes",
    "Camphene":"Terpinolene")
# view(AllTerpenesTable)

AllTerpenesTable <- pivot_longer(AllTerpenesTable,
  -c("Soil":"Total_Terpenes"),
  names_to = "Terpenes",
  values_to = "Values")

view(AllTerpenesTable)
##-----/Get Data 1-----

```

```

##-----Get Data 2 Bob & Mel-----

dataURL <-
  "https://docs.google.com/spreadsheets/d/1sup9_iIEcPW5ulxAtzm9BIDD8LWTANsl0sUKrM7l5s"

#query server for master sheet data
masterSheet <-
  read_sheet(dataURL,
    range = 'Master Sheet (corrected)!A3:CD45',
    .name_repair = unique) %>% janitor::clean_names(parsing_option=3)

#query server for soil info data
soilInfo <-
  read_sheet(dataURL,
    range = ('Soil_Info!A1:F6'),
    .name_repair = unique) %>% janitor::clean_names()

#create tibbles - blind for initial analysis

allthedata <- masterSheet

metaData <- masterSheet %>% select(soil:wet_weight_harvested)

plantCannabinoidTotal <-
  masterSheet %>% select(total_thc_plant:total_terpenes_plant)
view(plantCannabinoidTotal)

plantTissueElementalAnalysis <-
  masterSheet %>% select(total_nitrogen_plant:tissue_nitrate_nitrogen_plant)

elementRatios <-
  masterSheet %>% select(ca_mg_soil:mg_t_mg_s_plant_soil)

traceElements <-
  masterSheet %>% select(boron_ppm_soil:aluminum_ppm_soil)

percent <-
  masterSheet %>% select(calcium_percent_soil:sodium_percent_soil)

solubleCations <-
  masterSheet %>% select(calcium_ppm_soil:sodium_ppm_soil)

anions <-
  masterSheet %>% select(sulfur_ppm_soil:phosphorus_ppm_soil)

monoTerpenes <-
  masterSheet %>% select(camphene_plant:ocimene_isomer_l_plant)

sesquiTerpenes <-
  masterSheet %>% select(alpha_bisabolol_plant:cis_nerolidol_plant)

dirt <-
  masterSheet %>% select(p_h_soil:bicarbonate_hco3_ppm_soil)

```

```

#ID'd tibbles

IDMe <-
  masterSheet %>% select(planter_position, strain)

plantCannabinoidTotal <-
  masterSheet %>% select(planter_position, total_thc_plant:total_terpenes_plant)

IDplantTissueElementalAnalysis <-
  masterSheet %>% select(planter_position,
    total_nitrogen_plant:tissue_nitrate_nitrogen_plant)

IDelementRatios <-
  masterSheet %>% select(planter_position, ca_mg_soil:mg_t_mg_s_plant_soil)

IDtraceElements <-
  masterSheet %>% select(planter_position, boron_ppm_soil:aluminum_ppm_soil)

IDpercent <-
  masterSheet %>% select(planter_position,calcium_percent_soil:sodium_percent_soil)

IDSolubleCations <-
  masterSheet %>% select(planter_position,calcium_percent_soil:sodium_percent_soil)

IDanions <-
  masterSheet %>% select(planter_position,sulfur_ppm_soil:phosphorus_ppm_soil)

IDmonoTerpenes <-
  masterSheet %>% select(planter_position,camphene_plant:ocimene_isomer_l_plant)

IDSesquiTerpenes <-
  masterSheet %>% select(planter_position,
    alpha_bisabolol_plant:cis_nerolidol_plant)

IDdirt <-
  masterSheet %>% select(planter_position, p_h_soil:bicarbonate_hco3_ppm_soil)

##-----/ get data 2-----

##-----get data 3 (luke)-----
TableMaster <-
  masterSheet %>% select(total_terpenes_plant, total_nitrogen_plant,planter_position,
    soil_location, strain)

TableMaster2 <-
  masterSheet %>% select(total_terpenes_plant, total_thc_plant, planter_position,
    soil_location, strain)
##-----/get data 3-----

##-----Get data 4, Corrplots, (Mel & Anika)-----
#Percy told me about a package to play with so I did

```

```

IAMCurious <-
  allthedata %>% select(planter_position, p_t_plant:cis_nerolidol_plant)

IAMCurious3 <-
  allthedata %>% select(planter_position,wet_weight_harvested:cis_nerolidol_plant)

CuriousSoil <-
  masterSheet %>%
  select(strain, planter_position, total_terpenes_plant, total_thc_plant,
    total_nitrogen_plant,total_carbon_plant, p_h_soil:bicarbonate_hco3_ppm_soil,
    sulfur_ppm_soil:phosphorus_ppm_soil, calcium_ppm_soil:sodium_ppm_soil)

CuriousSoil$StID <- paste(CuriousSoil$strain, CuriousSoil$planter_position)

CuriousSoil2 <-
  CuriousSoil %>% select(StID, total_terpenes_plant:sodium_ppm_soil)
#filters by strain
FROG <- filter(CuriousSoil2, StID=="FR A3"
  | StID=="FR A4"
  | StID=="FR A6"
  | StID=="FR B3"
  | StID=="FR B4"
  | StID=="FR B6"
  | StID=="FR C3"
  | StID=="FR C4"
  | StID=="FR C6"
  | StID=="FR D3"
  | StID=="FR D4"
  | StID=="FR D6"
  | StID=="FR E3"
  | StID=="FR E4"
  | StID=="FR E6")

PHK <- filter(CuriousSoil2, StID=="PHK A1"
  | StID=="PHK A2"
  | StID=="PHK A5"
  | StID=="PHK B1"
  | StID=="PHK B2"
  | StID=="PHK B5"
  | StID=="PHK C1"
  | StID=="PHK C2"
  | StID=="PHK C5"
  | StID=="PHK D1"
  | StID=="PHK D2"
  | StID=="PHK D5"
  | StID=="PHK E1"
  | StID=="PHK E2"
  | StID=="PHK E5")

DIRTFROG <- filter(CuriousSoil2, StID=='FR G13'
  | StID=='FR G14'
  | StID=='FR G16'
  | StID=='FR G23'
  | StID=='FR G24'

```

```

|StID=='FR G26')

DIRTPHK <- filter(CuriousSoil2, StID=='PHK G11'
|StID=='PHK G12'
|StID=='PHK G15'
|StID=='PHK G21'
|StID=='PHK G22'
|StID=='PHK G25')

MonoTerps <-
  masterSheet %>%
  select(strain, planter_position, moisture_percent_plant:total_terpenes_plant,
    total_nitrogen_plant:total_carbon_plant,
    camphene_plant:ocimene_isomer_l_plant)
MonoTerps$StID <- paste(MonoTerps$strain, MonoTerps$planter_position)

MonoTerps2 <-
  MonoTerps %>% select(StID, moisture_percent_plant:ocimene_isomer_l_plant)

MFR0G <- filter(MonoTerps2, StID=="FR A3"
| StID=="FR A4"
| StID=="FR A6"
| StID=="FR B3"
| StID=="FR B4"
| StID=="FR B6"
| StID=="FR C3"
| StID=="FR C4"
| StID=="FR C6"
| StID=="FR D3"
| StID=="FR D4"
| StID=="FR D6"
| StID=="FR E3"
| StID=="FR E4"
| StID=="FR E6")

MPHK <- filter(MonoTerps2, StID=="PHK A1"
| StID=="PHK A2"
| StID=="PHK A5"
| StID=="PHK B1"
| StID=="PHK B2"
| StID=="PHK B5"
| StID=="PHK C1"
| StID=="PHK C2"
| StID=="PHK C5"
| StID=="PHK D1"
| StID=="PHK D2"
| StID=="PHK D5"
| StID=="PHK E1"
| StID=="PHK E2"
| StID=="PHK E5")

MDIRTFROG <- filter(MonoTerps2, StID=='FR G13'
|StID=='FR G14'

```

```

      | StID=='FR G16'
      | StID=='FR G23'
      | StID=='FR G24'
      | StID=='FR G26')

MDIRTPHK <- filter(MonoTerps2, StID=='PHK G11'
      | StID=='PHK G12'
      | StID=='PHK G15'
      | StID=='PHK G21'
      | StID=='PHK G22'
      | StID=='PHK G25')

MonoTerps3 <-
  masterSheet %>%
  select(moisture_percent_plant:total_terpenes_plant,
         total_nitrogen_plant:total_carbon_plant,
         camphene_plant:ocimene_isomer_l_plant)

sesquiTerp<-masterSheet %>%
  select(strain, planter_position, total_terpenes_plant,
         total_nitrogen_plant:total_carbon_plant,
         alpha_bisabolol_plant:cis_nerolidol_plant)

sesquiTerp$StID <- paste(sesquiTerp$strain, sesquiTerp$planter_position)

sesquiTerp2 <-
  sesquiTerp %>% select(StID, total_terpenes_plant:cis_nerolidol_plant)

SFR0G <- filter(sesquiTerp2, StID=="FR A3"
      | StID=="FR A4"
      | StID=="FR A6"
      | StID=="FR B3"
      | StID=="FR B4"
      | StID=="FR B6"
      | StID=="FR C3"
      | StID=="FR C4"
      | StID=="FR C6"
      | StID=="FR D3"
      | StID=="FR D4"
      | StID=="FR D6"
      | StID=="FR E3"
      | StID=="FR E4"
      | StID=="FR E6")

SPHK <- filter(sesquiTerp2, StID=="PHK A1"
      | StID=="PHK A2"
      | StID=="PHK A5"
      | StID=="PHK B1"
      | StID=="PHK B2"
      | StID=="PHK B5"
      | StID=="PHK C1"
      | StID=="PHK C2"
      | StID=="PHK C5")

```

```

      | StID=="PHK D1"
      | StID=="PHK D2"
      | StID=="PHK D5"
      | StID=="PHK E1"
      | StID=="PHK E2"
      | StID=="PHK E5")

SDIRTFROG <- filter(sesquiTerp2, StID=='FR G13'
      | StID=='FR G14'
      | StID=='FR G16'
      | StID=='FR G23'
      | StID=='FR G24'
      | StID=='FR G26')

SDIRTPHK <- filter(sesquiTerp2, StID=='PHK G11'
      | StID=='PHK G12'
      | StID=='PHK G15'
      | StID=='PHK G21'
      | StID=='PHK G22'
      | StID=='PHK G25')

dirtterp<-masterSheet%>%
  select(total_terpenes_plant,p_h_soil:bicarbonate_hco3_ppm_soil,
         sulfur_ppm_soil:phosphorus_ppm_soil,
         calcium_ppm_soil:sodium_ppm_soil,
         alpha_bisabolol_plant:cis_nerolidol_plant)

monodirt<-masterSheet%>%
  select(total_terpenes_plant,p_h_soil:bicarbonate_hco3_ppm_soil,
         sulfur_ppm_soil:phosphorus_ppm_soil,
         calcium_ppm_soil:sodium_ppm_soil,
         camphene_plant:ocimene_isomer_l_plant)

FROGMonoTerps <-
  MFRQG %>% select ( moisture_percent_plant:ocimene_isomer_l_plant)
#view(FROGMonoTerps)

PHKMonoTerps <-
  MPHKG %>% select ( moisture_percent_plant:ocimene_isomer_l_plant)

SFRQGTerps <-
  SFRQG %>% select(total_terpenes_plant,total_nitrogen_plant:total_carbon_plant,
                  alpha_bisabolol_plant:cis_nerolidol_plant)

SPHKTerps <-
  SPHKG %>% select (total_terpenes_plant,total_nitrogen_plant:total_carbon_plant,
                  alpha_bisabolol_plant:cis_nerolidol_plant)

MDirtTerpFROG <-
  MDIRTFROG %>% select ( moisture_percent_plant:ocimene_isomer_l_plant)

MDirtTerpPHK <-
  MDIRTPHK %>% select ( moisture_percent_plant:ocimene_isomer_l_plant)

SDirtTerpFROG <-
  SDIRTFROG %>% select(total_terpenes_plant,total_nitrogen_plant:total_carbon_plant,

```

```

                                alpha_bisabolol_plant:cis_nerolidol_plant)
SDirtTerpPKG <-
  SDIRTPHK %>% select(total_terpenes_plant,total_nitrogen_plant:total_carbon_plant,
                                alpha_bisabolol_plant:cis_nerolidol_plant)

sesquiTerp<-masterSheet %>%
  select(total_terpenes_plant,total_nitrogen_plant:total_carbon_plant,
                                alpha_bisabolol_plant:cis_nerolidol_plant)
dirtterp<-masterSheet%>%
  select(total_terpenes_plant,p_h_soil:bicarbonate_hco3_ppm_soil,
                                sulfur_ppm_soil:phosphorus_ppm_soil,
                                calcium_ppm_soil:sodium_ppm_soil,
                                alpha_bisabolol_plant:cis_nerolidol_plant)

monodirt<-masterSheet%>%
  select(total_terpenes_plant,p_h_soil:bicarbonate_hco3_ppm_soil,
                                sulfur_ppm_soil:phosphorus_ppm_soil,
                                calcium_ppm_soil:sodium_ppm_soil,camphene_plant:ocimene_isomer_l_plant)

IAMCurious2 <- IAMCurious %>% select(p_t_plant:cis_nerolidol_plant)
view(IAMCurious2)

IAMCurious4 <-
  allthedata %>% select(wet_weight_harvested:cis_nerolidol_plant)
##-----/get data 4-----

##-----corrplot matrices-----

M = cor(MonoTerps3)

O = cor(dirtterp)

K = cor(monodirt)

E = cor(IAMCurious2)
J = cor(IAMCurious4)

A = cor(FROGMonoTerps)
B = cor(PHKMonoTerps)
C = cor(SFROGTerps)
D = cor(SPHKTerps)
E = cor(MDirtTerpFROG)
G = cor(MDirtTerpPHK)
H = cor(SDirtTerpFROG)
I = cor(SDirtTerpPKG)
## -----end matrices-----

##plots

# The following ggplot() maps the following variables: Total Terpenes on the x axis, l
# individua terples on the y axis, and soil is mapped to colors according to the
# scale_color_viridis_c scale.
Ultisols_Terpenes <- AllTerpenesTable %>%
  filter(Soil_Order == "Ultisols")

```



```

Plot_Ultisols <- ggplot(Ultisols_Terpenes, aes(x = Total_Terpenes, y = Values)) +
  geom_point(aes(color = Terpenes)) +
  geom_line(aes(group = Terpenes, color = Terpenes)) +
  geom_label_repel(data = filter(Ultisols_Terpenes,
                                Total_Terpenes == max(Ultisols_Terpenes$Total_Terpenes)),
    aes(label = Terpenes)) +
  labs(x = "Total Terpene",
       y = "Individual Terpene",
       caption = "Figure 1. Total Terpene vs individual terpene values for Ultisols")+
  scale_fill_viridis_c()
#facet_wrap(~Soil_Subgroup)

Plot_Ultisols +
  theme(plot.caption = element_text(hjust = 0))
# theme(axis.title = element_text(family="Times", size = 10),
#       legend.title = element_text(family="Times", size = 10),
#       plot.caption = element_text(family="Times", size = 11, hjust = 0))
#Each Soil group
Mollisols_Terpenes <- AllTerpenesTable %>%
  filter(Soil_Order == "Mollisols")
view(Mollisols_Terpenes)

Plot_Mollisols <- ggplot(Mollisols_Terpenes, aes(x = Total_Terpenes, y = Values)) +
  geom_point(aes(color = Terpenes)) +
  geom_line(aes(group = Terpenes, color = Terpenes)) +
  geom_label_repel(data = filter(Mollisols_Terpenes,
                                Total_Terpenes == max(Mollisols_Terpenes$Total_Terpenes)),
    aes(label = Terpenes)) +
  scale_fill_viridis_c()+
  labs(x = "Total Terpene",
       y = "Individual Terpene",
       caption ="Figure 2. Total Terpene vs individual terpene values for Mollisols")

Plot_Mollisols

Alfisols_Terpenes <- AllTerpenesTable %>%
  filter(Soil_Order == "Alfisols")
view(Alfisols_Terpenes)

Plot_Alfisols <- ggplot(Alfisols_Terpenes, aes(x = Total_Terpenes, y = Values)) +
  geom_point(aes(color = Terpenes)) +
  geom_line(aes(group = Terpenes, color = Terpenes)) +
  geom_label_repel(data = filter(Alfisols_Terpenes,
                                Total_Terpenes == max(Alfisols_Terpenes$Total_Terpenes)),
    aes(label = Terpenes)) +
  scale_fill_viridis_c()+
  labs(x = "Total Terpene",
       y = "Individual Terpene",
       caption = " Figure 3. Total Terpene vs individual terpene values for Alfisols") +
  #geom_text(aes(label = Terpenes, size = NULL)) +
  facet_wrap(~Soil_Subgroup)
Plot_Alfisols

```

```

ggplot (data = TableMaster,
        mapping = aes(x = total_terpenes_plant, y = total_nitrogen_plant,
                      color= strain)) +
  geom_boxplot(aes(fill= strain)) +
  scale_colour_stata()+
  theme_stata()+
  stat_summary(fun=mean, geom="smooth", color="red") +
  labs( x = "Total Terpenes Plant (%)",
        y = "Total Nitrogen Plant (?)",
        caption= "Figure 4. Total Terpenes Plant vs. Total Nitrogen Plant (mean in red)")
set.seed(0)
t.test(TableMaster$total_terpenes_plant, TableMaster$total_nitrogen_plant,
        alternative = "two.sided", var.equal=FALSE)

ggplot (data = TableMaster2,
        mapping = aes(x = total_terpenes_plant, y= total_thc_plant,
                      color= soil_location, group_by(strain)))+
  geom_point()+
  geom_smooth(formula = y~x, method = "lm", se = TRUE)+
  labs( x = "Total Terpenes Plant (%)",
        y = "Total THC Plant (%)",
        caption= "Figure 5. Total Terpenes Plant vs. Total THC Plant") +
  scale_colour_stata()+
  theme_stata()+
  theme(axis.title = element_text(size = 10),
        legend.title = element_text(size = 10),
        plot.caption = element_text(size = 11, hjust = 0))

title <- 'Figure 6 - All data plotted in a correlation matrix'
corrplot(J, method='ellipse',na.label = ' ', tl.cex=0.5,
         title=title, mar=c(0,0,1,0))

testRes = cor.mtest(M, conf.level = 0.95)
corrplot(M, p.mat = testRes$p, method = 'number', insig='blank',
         number.cex = 0.5, order = 'AOE',
         diag=FALSE, tl.cex=0.7, pch.col = 'grey20',mar=c(0,0,1,0),
         title='Figure 7. Monoterpenes vs Total Carbon & Nitrogen, 95% CI' )

title2<-"Figure 8. Soil Nutrients vs. Sesquiterpenes"
testRes2 = cor.mtest(0, conf.level = 0.95)
corrplot(0, p.mat = testRes2$p, method = 'ellipse', type = 'lower', insig='blank',
         addCoef.col = 'black', number.cex = 0.5, order = 'AOE', diag=FALSE,
         tl.cex=0.7, title=title2, mar=c(0,0,1,0))

title3<-"Figure 9. Soil Nutrients vs. Monoterpenes"

testRes3 = cor.mtest(K, conf.level = 0.95)

corrplot(K, p.mat = testRes3$p, method = 'ellipse', type = 'lower', insig='blank',
         addCoef.col = 'black', number.cex = 0.5, order = 'AOE', diag=FALSE,
         tl.cex=0.7, title=title3, mar=c(0,0,1,0))

title4<-"Figure 10. FR vs. Monoterpenes"

```

```

testRes4 = cor.mtest(A, conf.level = 0.95)
corrplot(A, p.mat = testRes4$p, method = 'ellipse', insig='blank',
          number.cex = 0.5, order = 'hclust', diag=FALSE,
          addrect = 2, sig.level = 0.05, tl.cex=0.7, title=title4, mar=c(0,0,1,0))

title5<-"Figure 11. FR vs. Sesquiterpenes"
testRes5 = cor.mtest(C, conf.level = 0.95)
corrplot(C, p.mat = testRes5$p, method = 'ellipse', insig='blank',
          number.cex = 0.5, order = 'hclust', diag=FALSE,
          addrect = 2, sig.level = 0.05, tl.cex=0.7, title=title5, mar=c(0,0,1,0))

```