**2.6 Statistical analysis and plots- Powdery Mildew Scoring**

#PM Scoring

setwd(dir = "C:/Users/alish/OneDrive/Documents/Uni/MRes/Sanchez-Lucas\_Raw\_Datta\_et\_al")

#Load Libraries

library(Rmisc)

library(dplyr)

library(ggplot2)

library(ggpubr)

library(effectsize)

#Import Data and Clean

PM\_Data=read.csv('New\_PM\_Analysis.csv', header = T, sep = ',')

PM\_Data$YEAR=as.factor(PM\_Data$YEAR)

PM\_Data$MONTH=(PM\_Data$MONTH)

PM\_Data$ARRAY=as.factor(PM\_Data$ARRAY)

PM\_Data$TREE=as.factor(PM\_Data$TREE)

PM\_Data$CANOPY=as.factor(PM\_Data$CANOPY)

PM\_Data$CO2=as.factor(PM\_Data$CO2)

PM\_Data$Percentage.of.PM=as.numeric(PM\_Data$Percentage.of.PM)

PM\_Data=subset(PM\_Data, TREE!='UNKNOWN')

#Work out stats as per tree so the layers of the canopy are averaged

data\_as\_trees=PM\_Data%>%group\_by(CO2,YEAR,MONTH,TREE)%>%summarise(n=n(),mean=mean(Percentage.of.PM, na.rm=T), sd=sd(Percentage.of.PM))%>% mutate(se=sd/sqrt(n))

# Remove any unknowns

data\_as\_trees=na.omit(data\_as\_trees)

#Plotting all the Data

data\_as\_trees\_CO2\_Plot <- ggplot(PM\_Data, aes(x = CO2, y = Percentage.of.PM, fill = CO2)) +

geom\_boxplot() +

scale\_fill\_manual(values = c('dodgerblue3', 'orange1')) +

geom\_jitter(alpha = 0.5, position = position\_jitter(0.1)) +

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.5),

plot.title = element\_text(size = 15),

legend.position = 'none',

axis.text.y = element\_text(size = 15),

axis.text.x = element\_blank(),

axis.title.x = element\_blank(),

axis.title.y = element\_text(25),

axis.ticks.x = element\_blank()) +

ggtitle('All Data') +

ylab('% of PM') +

xlab(NULL) +

ylim(0, 100)

data\_as\_trees\_CO2\_Plot

#Stats on data\_as\_trees\_CO2\_Plot

ggqqplot(PM\_Data$Percentage.of.PM)

shapiro.test(PM\_Data$Percentage.of.PM)

p <- ggplot(PM\_Data, aes(x=Percentage.of.PM, color=CO2)) +

geom\_density()

p

ks.test(PM\_Data$Percentage.of.PM, 'pnorm')

wilcox.test(Percentage.of.PM ~ CO2, data =PM\_Data) # Not Significant

#Plotting Trees as Bioreplicates

data\_as\_replicates\_CO2\_Plot = ggplot(data\_as\_trees, aes(x = CO2, y = mean, fill = CO2)) +

geom\_boxplot() +

scale\_fill\_manual(values = c('dodgerblue3', 'orange1')) +

geom\_jitter(alpha = 0.5, position = position\_jitter(0.1)) +

ggtitle('Trees as Bioreplicates') +

ylab('% of PM') +

xlab('CO2 Treatment') +

ylim(0, 100) +

theme(

panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.5),

plot.title = element\_text(size = 15),

legend.position = 'none',

axis.text.y = element\_text(size = 15),

axis.text.x = element\_blank(),

axis.title.x = element\_blank(),

axis.title.y = element\_text(size = 15))

data\_as\_replicates\_CO2\_Plot

#Stats on data\_as\_replicates\_CO2\_Plot

ggqqplot(data\_as\_replicates\_CO2\_Plot$mean)

shapiro.test(data\_as\_replicates\_CO2\_Plot$mean)

p <- ggplot(data\_as\_replicates\_CO2\_Plot, aes(x=mean, color=CO2)) +

geom\_density()

p

wilcox.test(mean ~ CO2, data =data\_as\_replicates\_CO2\_Plot) # Not significant

cohens\_d(mean ~ CO2, data = data\_as\_trees) # Small effect size

#Look at differences in the canopy

data\_as\_trees\_canopy=PM\_Data %>% group\_by(TREE, CO2, YEAR, MONTH, CANOPY) %>% summarise(n=n(), mean=mean(Percentage.of.PM, na.rm=T), sd=sd(Percentage.of.PM)) %>% mutate(se=sd/sqrt(n))

#Separating this data into each canopy layer to plot

as\_trees\_top=subset(data\_as\_trees\_canopy, CANOPY=='Top')

as\_trees\_mid=subset(data\_as\_trees\_canopy, CANOPY=='Middle')

as\_trees\_bot=subset(data\_as\_trees\_canopy, CANOPY=='Bottom')

# Plot top data

as\_trees\_top\_plot <- ggplot(as\_trees\_top, aes(x = CO2, y = mean, fill = CO2)) +

geom\_boxplot() +

scale\_fill\_manual(values = c('dodgerblue3', 'orange1')) +

geom\_jitter(alpha = 0.2, position = position\_jitter(0.1)) +

ggtitle('Top Canopy') +

ylab('% of PM') +

ylim(0, 100) +

theme(

panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.5),

plot.title = element\_text(size = 20),

legend.position = 'none',

axis.text.y = element\_text(size = 20),

axis.title.y = element\_text(size=20),

axis.text.x = element\_blank(),

axis.title.x = element\_blank(),

axis.ticks.x = element\_blank())

as\_trees\_top\_plot

tiff("mature.tree.top.data.boxplot.tiff", units="cm", width=7.1, height=7.1, res=300)

as\_trees\_top\_plot

dev.off()

# Stats on top data

ggqqplot(as\_trees\_top$mean)

shapiro.test(as\_trees\_top$mean)

p <- ggplot(as\_trees\_top, aes(x=mean, color=CO2)) + geom\_density()

p

wilcox.test(mean ~ CO2, data =as\_trees\_top) # Not significant

cohens\_d(mean ~ CO2, data = as\_trees\_top) #Small effect size

#Plot middle data

as\_trees\_mid\_plot = ggplot(as\_trees\_mid, aes(x = CO2, y = mean, fill = CO2)) + geom\_boxplot() +

scale\_fill\_manual(values = c('dodgerblue3', 'orange1')) +

geom\_jitter(alpha = 0.2, position = position\_jitter(0.1)) +

ggtitle('Middle Canopy') + ylab('% of PM') + xlab('CO2 Treatment') + ylim(0, 100) + # This clips data outside the range

theme(panel.background = element\_blank(),axis.line = element\_line(color = 'black', size = 0.5),

plot.title = element\_text(size = 15),

legend.position = 'none',

axis.text.y = element\_text(size = 20),

axis.text.x = element\_blank(),

axis.title.x = element\_blank(),

axis.title.y = element\_text(size = 20))

as\_trees\_mid\_plot

#Stats on mid data

ggqqplot(as\_trees\_mid$mean)

shapiro.test(as\_trees\_mid$mean)

p <- ggplot(as\_trees\_mid, aes(x=mean, color=CO2)) +

geom\_density()

p

wilcox.test(mean ~ CO2, data =as\_trees\_mid) #Not significant

cohens\_d(mean ~ CO2, data = as\_trees\_mid) #Small effect size

#Plot bottom data

as\_trees\_bot\_plot = ggplot(as\_trees\_bot, aes(x = CO2, y = mean, fill = CO2)) +

geom\_boxplot() +

scale\_fill\_manual(values = c('dodgerblue3', 'orange1')) +

geom\_jitter(alpha = 0.2, position = position\_jitter(0.1)) +

ggtitle('Bottom Canopy') +

ylab('% of PM') +

xlab('CO2 Treatment') +

ylim(0, 100) +

theme(

panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.5),

plot.title = element\_text(size = 20),

legend.position = 'none',

axis.text.y = element\_text(size = 20),

axis.text.x = element\_blank(),

axis.title.x = element\_blank(),

axis.title.y = element\_text(size = 20))

as\_trees\_bot\_plot

#Stats on bottom

ggqqplot(as\_trees\_bot$mean)

shapiro.test(as\_trees\_bot$mean)

p <- ggplot(as\_trees\_bot, aes(x=mean, color=CO2)) +

geom\_density()

p

wilcox.test(mean ~ CO2, data =as\_trees\_bot) #Not significant

cohens\_d(mean ~ CO2, data = as\_trees\_bot) #Small effect

# CO2 over the year

co2overyear=data\_as\_trees %>% group\_by(CO2, MONTH) %>% summarise(n=n(), meanins=mean(mean), sd=sd(mean)) %>% mutate(se=sd/sqrt(n))

# CO2 over the years, month by month

co2overyear$MONTH <- factor(co2overyear$MONTH,

levels = c("May", "June", "July", "August", "September"),

ordered = TRUE)

co2overyear=na.omit(co2overyear)

CO2\_over\_year=ggplot(co2overyear, aes(x=MONTH, y=meanins, group=CO2, color=CO2))+

geom\_line()+

geom\_point()+

geom\_linerange(ymin=co2overyear$meanins-co2overyear$se, ymax=co2overyear$meanins+co2overyear$se)+

ggtitle('Average PM score % 2017-2024')+

ylab( '% of PM')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(), axis.line = element\_line(color = 'black',size = 0.5),plot.title = element\_text(size = 20),legend.position = 'none', axis.text.x = element\_text(size = 15), # X-axis tick label size

axis.text.y = element\_text(size = 15),

axis.title.x = element\_text(size = 18),

axis.title.y = element\_text(size = 18))

print(CO2\_over\_year)

#Stats on CO2\_over\_year

kruskal.test(meanins ~ MONTH, data =co2overyear) #No significant difference between months

co2btyeartree=data\_as\_trees %>% group\_by(CO2, MONTH, YEAR, TREE) %>% summarise(no=n(), meanins=mean(mean), std=sd(meanins)) %>% mutate(ste=std/sqrt(no))

co2btyear=co2btyeartree %>% group\_by(CO2, MONTH, YEAR) %>% summarise(n=n(), mean=mean(meanins), sd=sd(meanins)) %>% mutate(se=sd/sqrt(n))

as\_year\_2016=subset(co2btyear, YEAR=='2016')

as\_year\_2017=subset(co2btyear, YEAR=='2017')

as\_year\_2018=subset(co2btyear, YEAR=='2018')

as\_year\_2019=subset(co2btyear, YEAR=='2019')

as\_year\_2020=subset(co2btyear, YEAR=='2020')

as\_year\_2021=subset(co2btyear, YEAR=='2021')

as\_year\_2022=subset(co2btyear, YEAR=='2022')

as\_year\_2023=subset(co2btyear, YEAR=='2023')

as\_year\_2024=subset(co2btyear, YEAR=='2024')

#2016

as\_year\_2016\_plot <- ggplot(as\_year\_2016, aes(x = MONTH, y = mean, group = CO2, color = CO2)) +

geom\_point(size = 2) +

geom\_line(size = 1.5) +

geom\_linerange(aes(ymin = mean - se, ymax = mean + se), size = 1) +

scale\_color\_manual(values = c('dodgerblue3', 'grey')) +

labs(title = '2016', y = '% of PM', x = 'Month') +

theme\_minimal() +

theme(

panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.5),

plot.title = element\_text(size = 20),

axis.text.y = element\_text(size = 20),

axis.title.y = element\_text(size = 20),

axis.text.x = element\_text(size = 20),

axis.title.x = element\_blank(),

legend.position = 'none'

) +

ylim(25, 75)

as\_year\_2016\_plot

# Stats on 2016

wilcox.test(mean~CO2, as\_year\_2016) #Not significant

ggqqplot(as\_year\_2016\_plot$mean)

shapiro.test(as\_trees\_bot$mean)

p <- ggplot(as\_trees\_bot, aes(x=mean, color=CO2)) + geom\_density()

p

wilcox.test(mean ~ CO2, data =as\_trees\_bot) #Not significant

cohens\_d(mean ~ CO2, data = as\_trees\_bot)

#2017

as\_year\_2017$MONTH <- factor(as\_year\_2017$MONTH,

levels = c("August", "September"),ordered = TRUE)

as\_year\_2017=subset(as\_year\_2017, MONTH!='June')

as\_year\_2017\_plot=ggplot(as\_year\_2017, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2017$mean-as\_year\_2017$se, ymax=as\_year\_2017$mean+as\_year\_2017$se, size=1)+

ggtitle('2017')+

ylab('% of PM')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.5),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(25, 75)

as\_year\_2017\_plot

# Stats on as\_year\_2017\_plot

wilcox.test(mean~CO2, as\_year\_2017) #Not significant

#2018

as\_year\_2018$MONTH <- factor(as\_year\_2018$MONTH,

levels = c("June", "July", "August", "September"),

ordered = TRUE)

as\_year\_2018\_plot=ggplot(as\_year\_2018, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2018$mean-as\_year\_2018$se, ymax=as\_year\_2018$mean+as\_year\_2018$se, size=1)+

ggtitle('2018')+

ylab('% of PM')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.5),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(25, 75)

as\_year\_2018\_plot

# Stats on as\_year\_2018\_plot

wilcox.test(mean~CO2, as\_year\_2018) #Not significant

#2019

as\_year\_2019$MONTH <- factor(as\_year\_2019$MONTH,

levels = c("June", "July", "August", "September"),

ordered = TRUE)

as\_year\_2019\_plot=ggplot(as\_year\_2019, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2019$mean-as\_year\_2019$se, ymax=as\_year\_2019$mean+as\_year\_2019$se, size=1)+

ggtitle('2019')+

ylab('% of PM')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.5),

plot.title = element\_text(size = 18),

axis.text.y=element\_text(size=18),

axis.title.y = element\_text(size=18),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=18),

legend.position = 'none')+

ylim(25, 75)

as\_year\_2019\_plot

# Stats on as\_year\_2019\_plot

wilcox.test(mean~CO2, as\_year\_2019) #Not significant

#2020

as\_year\_2020$MONTH <- factor(as\_year\_2020$MONTH,

levels = c("June", "July", "August", "September"),

ordered = TRUE)

as\_year\_2020\_plot=ggplot(as\_year\_2020, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2020$mean-as\_year\_2020$se, ymax=as\_year\_2020$mean+as\_year\_2020$se, size=1)+

ggtitle('2020')+

ylab('% of PM')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.5),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(25, 75)

as\_year\_2020\_plot

# Stats on as\_year\_2020\_plot

wilcox.test(mean~CO2, as\_year\_2020) #Significance in September

#2021

as\_year\_2021$MONTH <- factor(as\_year\_2021$MONTH,

levels = c("June", "July","September"),

ordered = TRUE)

as\_year\_2021\_plot=ggplot(as\_year\_2021, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2021$mean-as\_year\_2021$se, ymax=as\_year\_2021$mean+as\_year\_2021$se, size=1)+

ggtitle('2021')+

ylab('% of PM')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.5),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(25, 75)

as\_year\_2021\_plot

# Stats on as\_year\_2021\_plot

wilcox.test(mean~CO2, as\_year\_2021)#not sif

#2022

as\_year\_2022$MONTH <- factor(as\_year\_2022$MONTH,

levels = c("May", "June", "July", "August", "September"),

ordered = TRUE)

as\_year\_2022\_plot <- ggplot(as\_year\_2022, aes(x = MONTH, y = mean, group = CO2, color = CO2)) +

geom\_point(size = 2) +

geom\_line(size = 1.5) +

geom\_linerange(aes(ymin = mean - se, ymax = mean + se), size = 1) +

ggtitle('2022') +

ylab('% of PM') +

xlab('Month') +

scale\_color\_manual(values = c('dodgerblue3', 'orange1')) +

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.5),

plot.title = element\_text(size = 25),

axis.text.y = element\_text(size = 25),

axis.title.y = element\_text(size = 25),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size = 25),

legend.position = 'none') +

ylim(25, 75)

as\_year\_2022\_plot

# Stats on as\_year\_2022\_plot

wilcox.test(mean~CO2, as\_year\_2022\_plot) #Not significant

#2023

as\_year\_2023$MONTH <- factor(as\_year\_2023$MONTH,

levels = c("May", "June", "July", "September"),

ordered = TRUE)

as\_year\_2023=na.omit(as\_year\_2023)

as\_year\_2023\_plot=ggplot(as\_year\_2023, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2023$mean-as\_year\_2023$se, ymax=as\_year\_2023$mean+as\_year\_2023$se, size=1)+

ggtitle('2023')+ylab("% of PM")+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.5),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=15),

legend.position = 'none')+

ylim(25, 75)

as\_year\_2023\_plot

# Stats on as\_year\_2023\_plot

wilcox.test(mean~CO2, as\_year\_2023) # Significance in May

#2024

as\_year\_2024$MONTH <- factor(as\_year\_2024$MONTH,

levels = c("May","July","September"),

ordered = TRUE)

as\_year\_2024\_plot=ggplot(as\_year\_2024, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2024$mean-as\_year\_2024$se, ymax=as\_year\_2024$mean+as\_year\_2024$se, size=1)+

ggtitle('2024')+

ylab("% of PM")+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.5),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(10, 75)

as\_year\_2024\_plot

# Stats on as\_year\_2024\_plot

wilcox.test(mean~CO2, as\_year\_2024)

# Data as arrays

data\_as\_trees\_array=PM\_Data %>% group\_by(CO2,ARRAY,TREE,YEAR,MONTH)%>%summarise(n=n(),mean=mean(Percentage.of.PM, na.rm=T), sd=sd(Percentage.of.PM))%>% mutate(se=sd/sqrt(n))

data\_as\_trees\_array=na.omit(data\_as\_trees\_array)

# Remove the year 2016

filtered\_data <- data\_as\_trees\_array %>% filter(YEAR != 2016)

# Create the plot

data\_as\_arrays\_Plot <- ggplot(filtered\_data, aes(x = ARRAY, y = mean, fill = ARRAY)) +

geom\_boxplot() +

scale\_fill\_manual(values = c('orange1', 'dodgerblue3', 'dodgerblue3', 'orange1', 'dodgerblue3', 'orange1')) +

geom\_jitter(alpha = 0.5, position = position\_jitter(0.1)) +

ylab('% of PM') +

ylim(25, 75) +

theme(

panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.5),

plot.title = element\_text(size = 25),

legend.position = 'none',

axis.text.y = element\_text(size = 25),

axis.text.x = element\_text(size = 25),

axis.title.x = element\_blank(),

axis.title.y = element\_text(size = 25))

data\_as\_arrays\_Plot

# Stats on data\_as\_arrays\_Plot

res\_aov <- aov(Percentage.of.PM ~ ARRAY,

data = PM\_Data)

summary(res\_aov)

kruskal.test(Percentage.of.PM ~ ARRAY,

data = PM\_Data)

# Significant differences

pairwise.wilcox.test(PM\_Data$Percentage.of.PM, PM\_Data$ARRAY,

p.adjust.method = "BH")

**2.6 Statistical analysis and plots- Insect Scoring**

#Insect Scoring

setwd(dir = "C:/Users/alish/OneDrive/Documents/Uni/YEAR 3/Lab Project")

#Load Libraries

library(Rmisc)

library(dplyr)

library(ggplot2)

library(ggpubr)

library(effectsize)

#Import Data and Clean

mature\_insect\_data=read.csv('insect scoring spreadsheet.xlsx - Sheet1 (2).csv', header = T, sep = ',')

mature\_insect\_data$YEAR=as.factor(mature\_insect\_data$YEAR)

mature\_insect\_data$MONTH=(mature\_insect\_data$MONTH)

mature\_insect\_data$ARRAY=as.factor(mature\_insect\_data$ARRAY)

mature\_insect\_data$TREE=as.factor(mature\_insect\_data$TREE)

mature\_insect\_data$CANOPY=as.factor(mature\_insect\_data$CANOPY)

mature\_insect\_data$CO2=as.factor(mature\_insect\_data$CO2)

mature\_insect\_data$LEAF.DAMAGE.INDEX=as.numeric(mature\_insect\_data$LEAF.DAMAGE.INDEX)

mature\_insect\_data=subset(mature\_insect\_data, YEAR!="2015")

mature\_insect\_data=subset(mature\_insect\_data, MONTH!='10')

mature\_insect\_data=subset(mature\_insect\_data, MONTH!='11')

mature\_insect\_data=subset(mature\_insect\_data, TREE!='UNKNOWN')

#Work out stats as per tree so the layers of the canopy are averaged

data\_as\_trees=mature\_insect\_data%>%group\_by(CO2,YEAR,MONTH,TREE)%>%summarise(n=n(),mean=mean(LEAF.DAMAGE.INDEX, na.rm=T), sd=sd(LEAF.DAMAGE.INDEX))%>% mutate(se=sd/sqrt(n))

data\_as\_trees=na.omit(data\_as\_trees)

#Plotting all Data

data\_as\_trees\_CO2\_Plot=ggplot(mature\_insect\_data, aes(x=CO2, y=LEAF.DAMAGE.INDEX, fill=CO2))+

geom\_boxplot()+

scale\_fill\_manual(values=c('dodgerblue3', 'orange1'))+

geom\_jitter(alpha=0.3, position = position\_jitter(0.1))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 12),

legend.position = 'none')+

ylab('Damage score %')+

ylim(0,100)+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 15),

axis.text.y=element\_text(size=15),

axis.text.x = element\_blank(),

axis.title.x = element\_blank(),

axis.title.y = element\_text(size=15),

axis.ticks.x = element\_blank())

data\_as\_trees\_CO2\_Plot

#Stats on data\_as\_trees\_CO2\_Plot

ggqqplot(mature\_insect\_data$LEAF.DAMAGE.INDEX)

shapiro.test(mature\_insect\_data$LEAF.DAMAGE.INDEX)

p <- ggplot(mature\_insect\_data, aes(x=LEAF.DAMAGE.INDEX, color=CO2)) +

geom\_density()

p

ks.test(mature\_insect\_data$LEAF.DAMAGE.INDEX, 'pnorm')

wilcox.test(LEAF.DAMAGE.INDEX ~ CO2, data =mature\_insect\_data) #Not significant

cohens\_d(mean ~ CO2, data = data\_as\_trees) #Small Effect size

#Data as Trees

data\_as\_bioreplicates\_CO2\_Plot=ggplot(data\_as\_trees, aes(x=CO2, y=mean, fill=CO2))+

geom\_boxplot()+

scale\_fill\_manual(values=c('dodgerblue3', 'orange1'))+

geom\_jitter(alpha=0.3, position = position\_jitter(0.1))+

theme(panel.background = element\_blank(), axis.line = element\_line(color = 'black',size = 0.3),plot.title = element\_text(size = 12), legend.position = 'none')+

ylab('Damage Score %')+

ylim(0,100)+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 15),

axis.text.y=element\_text(size=15),

axis.text.x = element\_blank(),

axis.title.x = element\_blank(),

axis.title.y = element\_text(size=15))

data\_as\_bioreplicates\_CO2\_Plot

# Stats on data\_as\_bioreplicates\_CO2\_Plot

ggqqplot(data\_as\_trees$mean)

shapiro.test(data\_as\_trees$mean)

p <- ggplot(data\_as\_trees, aes(x=mean, color=CO2)) +

geom\_density()

p

wilcox.test(mean ~ CO2, data =data\_as\_trees) #Not significant

cohens\_d(mean ~ CO2, data = data\_as\_trees) #Small effect size

#Look at Differences in the Canopy

data\_as\_trees\_canopy=mature\_insect\_data %>% group\_by(TREE, CO2, YEAR, MONTH, CANOPY) %>% summarise(n=n(), mean=mean(LEAF.DAMAGE.INDEX, na.rm=T), sd=sd(LEAF.DAMAGE.INDEX)) %>% mutate(se=sd/sqrt(n))

#Separating this data into each canopy layer to plot

as\_trees\_top=subset(data\_as\_trees\_canopy, CANOPY=='Top')

as\_trees\_mid=subset(data\_as\_trees\_canopy, CANOPY=='Middle')

as\_trees\_bot=subset(data\_as\_trees\_canopy, CANOPY=='Bottom')

#Plot Top data

as\_trees\_top\_plot = ggplot(as\_trees\_top, aes(x = CO2, y = mean, fill = CO2)) +

geom\_boxplot() +

scale\_fill\_manual(values = c('dodgerblue3', 'orange1')) +

geom\_jitter(alpha = 0.2, position = position\_jitter(0.1)) +

ggtitle('Top Canopy') +

ylab('Damage Score %') +

ylim(0, 100) +

theme(

panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.3),

plot.title = element\_text(size = 20),

legend.position = 'none',

axis.text.y = element\_text(size = 20),

axis.text.x = element\_blank(), # X-axis text labels removed

axis.title.x = element\_blank(), # X-axis title removed

axis.title.y = element\_text(size = 20))

as\_trees\_top\_plot

#Stats on as\_trees\_top\_plot

ggqqplot(as\_trees\_top$mean)

shapiro.test(as\_trees\_top$mean)

p <- ggplot(as\_trees\_top, aes(x=mean, color=CO2)) +

geom\_density()

p

wilcox.test(mean ~ CO2, data =as\_trees\_top) #Not significant

cohens\_d(mean ~ CO2, data = as\_trees\_top) #Small effect size

#Plot mid data

as\_trees\_mid\_plot = ggplot(as\_trees\_mid, aes(x = CO2, y = mean, fill = CO2)) +

geom\_boxplot() +

scale\_fill\_manual(values = c('dodgerblue3', 'orange1')) +

geom\_jitter(alpha = 0.2, position = position\_jitter(0.1)) +

ggtitle('Middle Canopy') +

ylab('Damage Score %') +

ylim(0, 100) + # Clips data outside range; consider coord\_cartesian if needed

theme(

panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.3),

plot.title = element\_text(size = 20),

legend.position = 'none',

axis.text.y = element\_text(size = 20),

axis.text.x = element\_blank(), # X-axis text labels removed

axis.title.x = element\_blank(), # X-axis title removed

axis.title.y = element\_text(size = 20))

as\_trees\_mid\_plot

#Stats on mid data

ggqqplot(as\_trees\_mid$mean)

shapiro.test(as\_trees\_mid$mean)

p <- ggplot(as\_trees\_mid, aes(x=mean, color=CO2)) +

geom\_density()

p

wilcox.test(mean ~ CO2, data =as\_trees\_mid) #Not significant

cohens\_d(mean ~ CO2, data = as\_trees\_mid) #Small effect size

#Plot bottom data

as\_trees\_bot\_plot = ggplot(as\_trees\_bot, aes(x = CO2, y = mean, fill = CO2)) +

geom\_boxplot() +

scale\_fill\_manual(values = c('dodgerblue3', 'orange1')) +

geom\_jitter(alpha = 0.2, position = position\_jitter(0.1)) +

ylab('Damage Score %') +

ggtitle('Bottom Canopy') +

ylim(0, 100) +

theme(

panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.3),

plot.title = element\_text(size = 20),

legend.position = 'none',

axis.text.y = element\_text(size = 20),

axis.text.x = element\_blank(),

axis.title.x = element\_blank(),

axis.title.y = element\_text(size = 20))

as\_trees\_bot\_plot

#Stats on as\_trees\_bot\_plot

ggqqplot(as\_trees\_bot$mean)

shapiro.test(as\_trees\_bot$mean)

p <- ggplot(as\_trees\_bot, aes(x=mean, color=CO2)) +

geom\_density()

p

wilcox.test(mean ~ CO2, data =as\_trees\_bot) #Not significant

cohens\_d(mean ~ CO2, data = as\_trees\_bot) #Small effect

#Look at CO2 over year

co2overyear = data\_as\_trees %>%

group\_by(CO2, MONTH) %>%

summarise(n = n(), meanins = mean(mean), sd = sd(mean)) %>%

mutate(se = sd / sqrt(n))

# Order the months as a factor

co2overyear$MONTH <- factor(co2overyear$MONTH,

levels = c("May", "June", "July", "August", "September"),

ordered = TRUE)

co2overyear = na.omit(co2overyear)

# Create the plot

CO2\_over\_year = ggplot(co2overyear, aes(x = MONTH, y = meanins, group = CO2, color = CO2)) +

geom\_line() +

geom\_point() +

geom\_linerange(aes(ymin = meanins - se, ymax = meanins + se)) +

ggtitle('Average Damage score % 2017-2024') +

ylab('Damage Score %') +

xlab('Month') +

scale\_color\_manual(values = c('dodgerblue3', 'orange1')) +

theme(

panel.background = element\_blank(),

axis.line = element\_line(color = 'black', size = 0.5),

plot.title = element\_text(size = 20),

legend.position = 'none',

axis.text.x = element\_text(size = 15),

axis.text.y = element\_text(size = 15),

axis.title.x = element\_text(size = 18),

axis.title.y = element\_text(size = 18)

) +

coord\_cartesian(ylim = c(20, 40))

CO2\_over\_year

#Stats on CO2\_over\_year

kruskal.test(meanins ~ MONTH, data =co2overyear) #Not significant

co2btyeartree=data\_as\_trees %>% group\_by(CO2, MONTH, YEAR, TREE) %>% summarise(no=n(), meanins=mean(mean), std=sd(meanins)) %>% mutate(ste=std/sqrt(no))

co2btyear=co2btyeartree %>% group\_by(CO2, MONTH, YEAR) %>% summarise(n=n(), mean=mean(meanins), sd=sd(meanins)) %>% mutate(se=sd/sqrt(n))

as\_year\_2016=subset(co2btyear, YEAR=='2016')

as\_year\_2017=subset(co2btyear, YEAR=='2017')

as\_year\_2018=subset(co2btyear, YEAR=='2018')

as\_year\_2019=subset(co2btyear, YEAR=='2019')

as\_year\_2020=subset(co2btyear, YEAR=='2020')

as\_year\_2021=subset(co2btyear, YEAR=='2021')

as\_year\_2022=subset(co2btyear, YEAR=='2022')

as\_year\_2023=subset(co2btyear, YEAR=='2023')

as\_year\_2024=subset(co2btyear, YEAR=='2024')

#2016

as\_year\_2016$MONTH <- factor(as\_year\_2016$MONTH,

levels = c("June", "July", "August", "September"),

ordered = TRUE)

# Remove rows with NAs

as\_year\_2016 <- na.omit(as\_year\_2016)

as\_year\_2016\_plot=ggplot(as\_year\_2016, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2016$mean-as\_year\_2016$se, ymax=as\_year\_2016$mean+as\_year\_2016$se, size=1)+

ggtitle('2016')+

ylab('Damage Score %')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'grey'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 25),

axis.text.y=element\_text(size=25),

axis.title.y = element\_text(size=25),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=25),

legend.position = 'none')+

ylim(10, 60)

as\_year\_2016\_plot

# Stats on as\_year\_2016\_plot

wilcox.test(mean~CO2, as\_year\_2016) #Not significant

#2017

as\_year\_2017$MONTH <- factor(as\_year\_2017$MONTH,

levels = c("June", "July", "August", "September"),

ordered = TRUE)

as\_year\_2017\_plot=ggplot(as\_year\_2017, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2017$mean-as\_year\_2017$se, ymax=as\_year\_2017$mean+as\_year\_2017$se, size=1)+

ggtitle('2017')+

ylab('Damage Score %')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 25),

axis.text.y=element\_text(size=25),

axis.title.y = element\_text(size=25),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=25),

legend.position = 'none')+

ylim(10, 60)

as\_year\_2017\_plot

# Stats on as\_year\_2017\_plot

wilcox.test(mean~CO2, as\_year\_2017\_plot) #Not significant

#2018

as\_year\_2018$MONTH <- factor(as\_year\_2018$MONTH,

levels = c("June", "July", "August", "September"),

ordered = TRUE)

as\_year\_2018\_plot=ggplot(as\_year\_2018, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2018$mean-as\_year\_2018$se, ymax=as\_year\_2018$mean+as\_year\_2018$se, size=1)+

ggtitle('2018')+

ylab('Damage Score %')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(10, 60)

as\_year\_2018\_plot

# Stats on as\_year\_2018\_plot

wilcox.test(mean~CO2, as\_year\_2018\_plot) #Not significant

#2019

as\_year\_2019$MONTH <- factor(as\_year\_2019$MONTH,

levels = c("June", "July", "August", "September"),

ordered = TRUE)

as\_year\_2019\_plot=ggplot(as\_year\_2019, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2019$mean-as\_year\_2019$se, ymax=as\_year\_2019$mean+as\_year\_2019$se, size=1)+

ggtitle('2019')+

ylab('Damage Score %')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(10, 60)

as\_year\_2019\_plot

# Stats on as\_year\_2019\_plot

wilcox.test(mean~CO2, as\_year\_2019) #Not significant

#2020

as\_year\_2020$MONTH <- factor(as\_year\_2020$MONTH,

levels = c("June", "July", "August", "September"),

ordered = TRUE)

as\_year\_2020= na.omit(as\_year\_2020)

as\_year\_2020\_plot=ggplot(as\_year\_2020, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2020$mean-as\_year\_2020$se, ymax=as\_year\_2020$mean+as\_year\_2020$se, size=1)+

ggtitle('2020')+

ylab('Damage Score %')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(10, 60)

as\_year\_2020\_plot

# Stats on as\_year\_2020\_plot

wilcox.test(mean~CO2, as\_year\_2020) #Not significant

#2021

as\_year\_2021$MONTH <- factor(as\_year\_2021$MONTH,

levels = c("June", "July", "August", "September"),

ordered = TRUE)

as\_year\_2021\_plot=ggplot(as\_year\_2021, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2021$mean-as\_year\_2021$se, ymax=as\_year\_2021$mean+as\_year\_2021$se, size=1)+

ggtitle('2021')+

ylab('Damage Score %')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(10, 60)

as\_year\_2021\_plot

# Stats on as\_year\_2021\_plot

wilcox.test(mean~CO2, as\_year\_2021) #Not significant

#2022

as\_year\_2022$MONTH <- factor(as\_year\_2022$MONTH,

levels = c("May", "June", "July", "August", "September"),

ordered = TRUE)

as\_year\_2022\_plot=ggplot(as\_year\_2022, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2022$mean-as\_year\_2022$se, ymax=as\_year\_2022$mean+as\_year\_2022$se, size=1)+

ggtitle('2022')+

ylab('Damage Score %')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 25),

axis.text.y=element\_text(size=25),

axis.title.y = element\_text(size=25),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=25),

legend.position = 'none')+

ylim(20, 60)

as\_year\_2022\_plot

# Stats on as\_year\_2022\_plot

wilcox.test(mean~CO2, as\_year\_2022) #Not significant

#2023

as\_year\_2023$MONTH <- factor(as\_year\_2023$MONTH,

levels = c("May", "June", "July"),

ordered = TRUE)

as\_year\_2023=na.omit(as\_year\_2023)

as\_year\_2023\_plot=ggplot(as\_year\_2023, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2023$mean-as\_year\_2023$se, ymax=as\_year\_2023$mean+as\_year\_2023$se, size=1)+

ggtitle('2023')+

ylab('Damage Score %')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(10, 60)

as\_year\_2023\_plot

# Stats on as\_year\_2023\_plot

wilcox.test(mean~CO2, as\_year\_2023)

#2024

as\_year\_2024$MONTH <- factor(as\_year\_2024$MONTH,

levels = c("May", "July", "September"),

ordered = TRUE)

as\_year\_2024=na.omit(as\_year\_2024)

as\_year\_2024\_plot=ggplot(as\_year\_2024, aes(x=MONTH, y=mean, group=CO2, color=CO2))+

geom\_point(size=2)+

geom\_line(size=1.5)+

geom\_linerange(ymin=as\_year\_2024$mean-as\_year\_2024$se, ymax=as\_year\_2024$mean+as\_year\_2024$se, size=1)+

ggtitle('2024')+

ylab('Damage Score %')+

xlab('Month')+

scale\_color\_manual(values=c('dodgerblue3', 'orange1'))+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 20),

axis.text.y=element\_text(size=20),

axis.title.y = element\_text(size=20),

axis.title.x = element\_blank(),

axis.text.x = element\_text(size=20),

legend.position = 'none')+

ylim(10, 60)

as\_year\_2024\_plot

# Stats on as\_year\_2024\_plot

wilcox.test(mean~CO2, as\_year\_2024) #Not significant

# Plotting data as arrays

data\_as\_trees\_array=mature\_insect\_data %>% group\_by(CO2,ARRAY,TREE,YEAR,MONTH)%>%summarise(n=n(),mean=mean(LEAF.DAMAGE.INDEX, na.rm=T), sd=sd(LEAF.DAMAGE.INDEX))%>% mutate(se=sd/sqrt(n))

data\_as\_trees\_array=na.omit(data\_as\_trees\_array)

data\_as\_arrays\_Plot=ggplot(data\_as\_trees\_array, aes(x=ARRAY, y=mean, fill=ARRAY))+

geom\_boxplot()+

scale\_fill\_manual(values=c('orange1', 'dodgerblue3','dodgerblue3','orange1','dodgerblue3','orange1'))+

geom\_jitter(alpha=0.3, position = position\_jitter(0.1))+

theme(panel.background = element\_blank(), axis.line = element\_line(color = 'black',size = 0.3),plot.title = element\_text(size = 12), legend.position = 'none')+

ylab('Damage score %')+

ylim(0,100)+

theme(panel.background = element\_blank(),

axis.line = element\_line(color = 'black',size = 0.3),

plot.title = element\_text(size = 15),

axis.text.y=element\_text(size=15),

axis.text.x = element\_text(size=15),

axis.title.x = element\_blank(),

axis.title.y = element\_text(size=15))

data\_as\_arrays\_Plot

#Stats on data\_as\_arrays\_Plot

res\_aov <- aov(LEAF.DAMAGE.INDEX ~ ARRAY,

data = mature\_insect\_data

)

summary(res\_aov)

kruskal.test(LEAF.DAMAGE.INDEX ~ ARRAY,

data = mature\_insect\_data)

# Significant differences

pairwise.wilcox.test(mature\_insect\_data$LEAF.DAMAGE.INDEX, mature\_insect\_data$ARRAY,

p.adjust.method = "BH")