# Independent Project Introduction

Savannah K. Miller

2025-03-09

**Dataset Card Source: (https://github.com/Smillersfsu/DepositionSFB/blob/0a6b8b2033b76344b3b1574f88ab81aa244e5210/DatasetCard.md)**

## Background

Sediment is the structural foundation of the salt marsh ecosystem where endemic species live and acquire resources(Levin et al. 2001; Janousek et al. 2017). The salt marsh ecosystem is host to many endemic species that are specially adapted to the conditions of the salt marsh(Dunson and Travis 1994; Schile et al. 2011; Moffett, Robinson, and Gorelick 2010; Rankin et al. 2023), and the very sediment they walk and live within is in short supply(Schoellhamer and Marineau, n.d.). Salt marshes play strong roles in dampening impacts of atmospheric rivers and extreme weather events, offering a protection to coastal communities(Smolders et al. 2015; Castagno et al. 2022; Lee and Nepf 2024; Taylor-Burns et al. 2024). Already, there has been a long human history of disturbance to existing salt marshes, which puts ecological and human communities at severe risk(Gedan, Silliman, and Bertness 2009). Under these circumstances, the few salt marshes left are at risk of elimination due to drowning if action is not taken and the benefits of this ecosystem could be forever lost.(Endris et al. 2024).

Across the state, salt marshes are decreasing in number and area due to a variety of factors(Endris et al. 2024). This unfortunate circumstance paired with changing water levels along the coastline increases the sense of urgency to understand, conserve, and restore California's salt marshes. To assist in raising the elevation of the salt marshes to adapt to natural hazards, we need to understand the fundamental processes of how sediment accumulates within the salt marsh over a brief period (short-term deposition) (Fagherazzi 2013; Houttuijn Bloemendaal et al. 2021; Vandenbruwaene et al. 2011). In this study, I will measure short-term deposition rates across marsh edges of multiple marshes within a single estuary in San Pablo Bay of the San Francisco Bay Estuary and comparing those rates to measurements of the surrounding marsh landscape. This study will provide information to ongoing restoration efforts of salt marshes in San Pablo Bay.(Callaway 2024; University of San Francisco et al. 2011; K. M. Thorne et al. 2019; Haltiner et al. 1996) My study provides novel insights for land managers to characterize their marshes edges to adapt to natural hazards and enhance community resiliency(Reed et al. 1999).

In this study, I will measure short-term deposition rates on multiple marshes within the San Francisco Bay Estuary and will compare those rates to measurements of the surrounding marsh landscape. This study builds on the work of Karen Thorne and U.S. Geological Survey (2025; Inter-and Intra-Annual Sediment Dynamics in Two Tidal Marshes: Deposition, Accretion, and Vegetation Data - ScienceBase-Catalog) to understand differences in sediment deposition on marshes with different edge characteristics, vegetation densities and geographic settings. Moreover, I will measure variables such as elevation, vegetation structure, and observations of marsh edge traits and comparing to rates of short-term deposition measured to understand patterns between variables. Other environmental variables such as season at time of collection and tide level at time of collection will be recorded and considered in analyses.

### Research Objective

This study will measure short-term deposition rates of sediment across the exposed (bay-adjacent) salt marshes in the northern part of the San Francisco Bay Estuary of the San Pablo Bay. I will compare these rates to biological structures of each salt marsh.

### Research Aim:

Marsh characteristics of the vegetation also are one of the first interactions with the waves bay ward and along the marsh edge. These vegetation characteristics will be measured at each marsh and compared with the rates of sediment deposition measured over the course of a year's spring-neap tidal cycles.

**Expected Outcome:** Deposition rates vary significantly with characteristics of the marsh vegetation, specifically maximum height.

## Null Hypothesis:

Deposition rates of sediment do not vary with differences in height of Salicornia pacificia (Pickleweed).

## Alternate hypothesis

Deposition rates of sediment do vary with differences in height of Salicornia pacifica (Pickleweed).

### Data analysis Plan

First, with steps 1- , I played with both the vegetation and sediment datasets to understand patterns that each of those explanatory variables had with the landscape and abiotic factors, such as distance from the water source ("Distance"), weather factors ("Season"), time factors ("Year" and "Month"), and location ("Site").

After going through the steps, I went through a two sampled t-test to see if there was a relationship between the sediment fluxes, and my hypothesized factor of maximum plant height of Salicornia pacifica versus the location of the measurements ("Site") to see if those explanatory variables were different between the two sites.

I then ran a general linear mixed model with both sites data and sediment deposition versus the maximum plant height. I followed this by running a general linear mixed model for each site with those same explanatory variables to compare the AIC values and find the best fit. The general linear mixed models I believe demonstrate from their values the clearest relationship that the sites have different relationships with the variables due to Corte Madera's much lower value than San Pablo's glmm, despite them both having the same factors accounted for in the model.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.4     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(Rmisc)
```

```
## Warning: package 'Rmisc' was built under R version 4.4.3
```

```
## Loading required package: lattice
## Loading required package: plyr
## ------------------------------------------------------------------------------
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## ------------------------------------------------------------------------------
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
##
## The following object is masked from 'package:purrr':
##
##     compact
```

```r
library(rstatix)
```

```
## Warning: package 'rstatix' was built under R version 4.4.3
```

```
##
## Attaching package: 'rstatix'
##
## The following objects are masked from 'package:plyr':
##
##     desc, mutate
##
## The following object is masked from 'package:stats':
##
##     filter
```

```r
library(tidyr)
library(data.table)
```

```
## Warning: package 'data.table' was built under R version 4.4.3
```

```
##
## Attaching package: 'data.table'
##
## The following objects are masked from 'package:lubridate':
##
##      hour, isoweek, mday, minute, month, quarter, second, wday, week,
##      yday, year
##
## The following objects are masked from 'package:dplyr':
##
##      between, first, last
##
## The following object is masked from 'package:purrr':
##
##      transpose
```

```r
library(dplyr)
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 4.4.3
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##      expand, pack, unpack
```

```r
library(stringr)
```

```r
knitr::opts_knit$set(root.dir = 'C:/Users/savan/OneDrive/Documents/Code/GitHub/DepositionSFB/data')
```

```r
# load data
dataset_sed <- read.csv("C:/Users/savan/OneDrive/Documents/Code/GitHub/DepositionSFB/data/SFB_sediment_

dataset_veg <- read.csv("C:/Users/savan/OneDrive/Documents/Code/GitHub/DepositionSFB/data/SFB_1m_vegeta

sed <- dataset_sed %>% mutate(FluxAvgRep = ifelse(is.na(FluxAvgRep), 0, FluxAvgRep))

sed <- dataset_sed %>%
  select(-Date.removed.from.field, -Date.placed.in.field)

sed <- sed %>% mutate(FluxAvgRep = ifelse(is.na(FluxAvgRep), 0, FluxAvgRep))

#sed$Plot.ID <- paste(sed$Transect, sed$Distance)
#veg$Plot.ID <- paste(veg$Transect, veg$Distance)
sed$Season <- str_replace_all(sed$Season, " ", "")
sed$Distance <- str_replace_all(sed$Distance, " ", "")
sed$Transect <- str_replace_all(sed$Transect, " ", "")
sed$Year <- str_replace_all(sed$Year, " ", "")
sed$Month <- str_replace_all(sed$Month, " ", "")
```

```
sed$FluxAvgRep <- str_replace_all(sed$FluxAvgRep, " ", "")
#sed$Plot.ID <- str_replace_all(sed$Plot.ID, " ", "")
sed$FluxAvgRep <-as.numeric(sed$FluxAvgRep)
```

```
## Warning: NAs introduced by coercion
```

```
sed$FluxAvgRep <- round(sed$FluxAvgRep)
sed$Year <- as.character(sed$Year)
sed$Distance <-as.numeric(sed$Distance)


veg <- dataset_veg %>% mutate(Avg.Ht = ifelse(is.na(Avg.Ht), 0, Avg.Ht))
veg <- dataset_veg %>% mutate(Max.Ht = ifelse(is.na(Max.Ht), 0, Max.Ht))
veg$Season <- str_replace_all(veg$Season, " ", "")
veg$Distance <- str_replace_all(veg$Distance, " ", "")
veg$Transect <- str_replace_all(veg$Transect, " ", "")
veg$Year <- str_replace_all(veg$Year, " ", "")
veg$Month <- str_replace_all(veg$Month, " ", "")
veg$Year <- as.character(veg$Year)
veg$Cover <-as.numeric(veg$Cover)
veg$Distance <-as.numeric(veg$Distance)

str(sed)
```

```
## 'data.frame':    183 obs. of  7 variables:
##  $ Site      : chr  "Corte Madera" "Corte Madera" "Corte Madera" "Corte Madera" ...
##  $ Transect  : chr  "A" "A" "A" "A" ...
##  $ Distance  : num  2 6 12 24 48 0.5 2 6 12 24 ...
##  $ Month     : chr  "AUGUST" "AUGUST" "AUGUST" "AUGUST" ...
##  $ Season    : chr  "Summer" "Summer" "Summer" "Summer" ...
##  $ Year      : chr  "2022" "2022" "2022" "2022" ...
##  $ FluxAvgRep: num  4 2 8 18 6 0 3 4 8 24 ...
```

```
str(veg)
```

```
## 'data.frame':    115 obs. of  11 variables:
##  $ Month   : chr  "AUGUST" "AUGUST" "AUGUST" "AUGUST" ...
##  $ Season  : chr  "Summer" "Summer" "Summer" "Summer" ...
##  $ Year    : chr  "2022" "2022" "2022" "2022" ...
##  $ Date    : chr  "8/24/2022" "8/24/2022" "8/24/2022" "8/24/2022" ...
##  $ Site    : chr  "Corte Madera" "Corte Madera" "Corte Madera" "Corte Madera" ...
##  $ Transect: chr  "A" "A" "A" "A" ...
##  $ Distance: num  0.5 12 2 24 6 0.5 12 24 48 6 ...
##  $ Species : chr  "SAPA" "SAPA" "SAPA" "SAPA" ...
##  $ Avg.Ht  : num  44 28 26 30 31 26 28 27 21 22 ...
##  $ Max.Ht  : num  54 41 37 45 39 50 35 37 37 38 ...
##  $ Cover   : num  60 10 30 50 5 55 45 15 50 30 ...
```

```
sed_mean <- mean(sed$FluxAvgRep, na.rm = TRUE)
sed_mean
```

```
## [1] 15.91758
```

```
sed_max  <- max(sed$FluxAvgRep, na.rm = TRUE)
sed_max
```
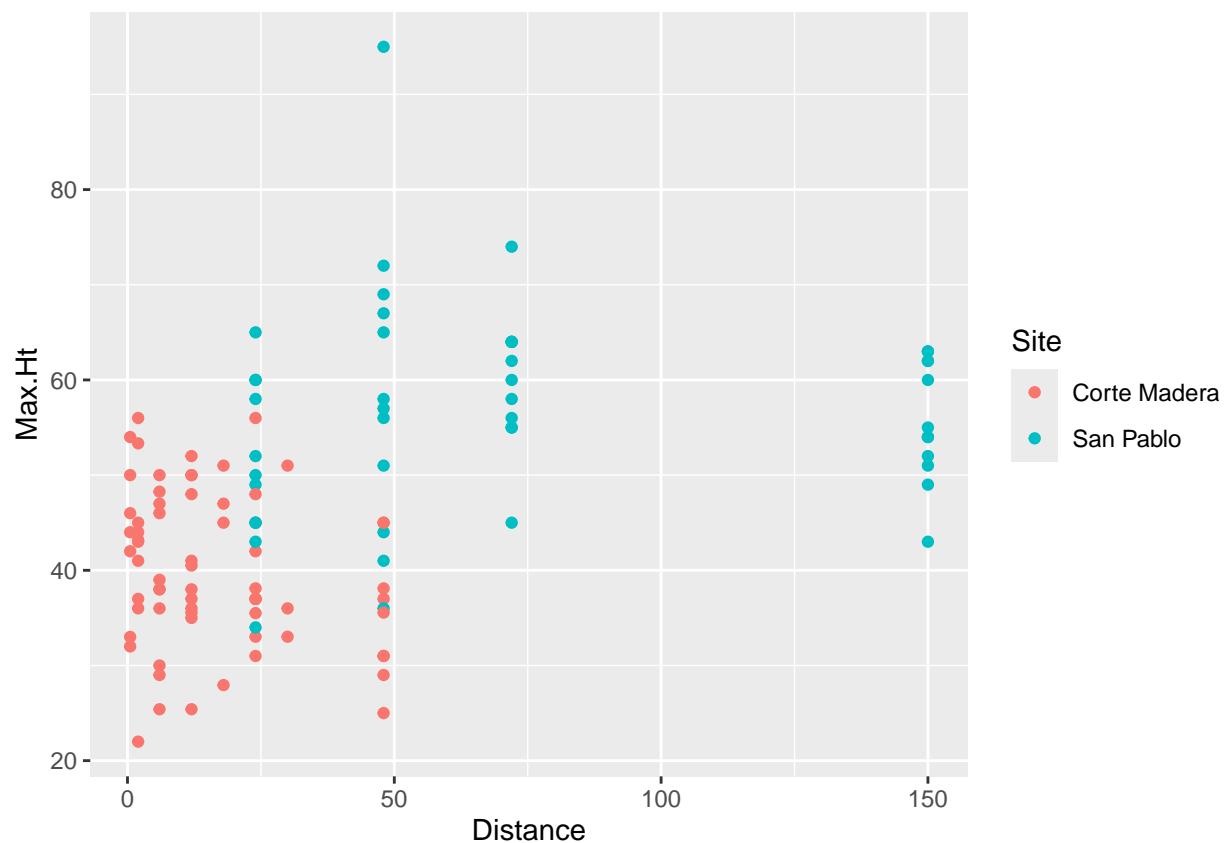
```
## [1] 356
```

```
sed_min  <- min(sed$FluxAvgRep, na.rm = TRUE)
sed_min
```

```
## [1] 0
```

```
sed_sd <-sd(sed$FluxAvgRep, na.rm =TRUE)
sed_sd
```

```
## [1] 41.69448
```

```
SAPA_time_space <- ggplot(data=veg, aes(x=Distance, y=Max.Ht, color=Site)) + geom_point()
SAPA_time_space
```



```
sed_space <-ggplot(data=sed, aes(x=Distance, y=FluxAvgRep, color=Site)) +geom_point() + geom_jitter(wid
 coord_cartesian(ylim = c(0, 357))
sed_space
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_point()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_point()').
```
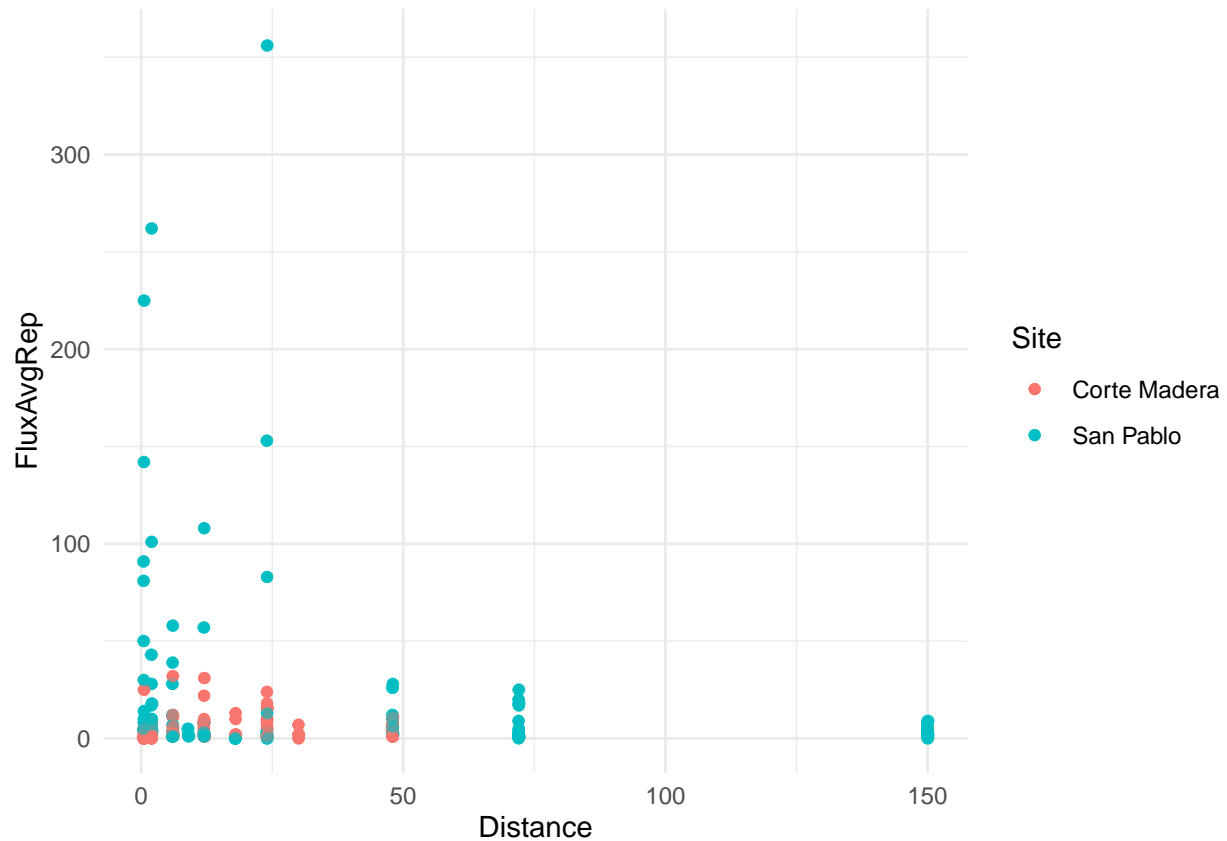


```
sed_time <-ggplot(data=sed, aes(x=Season, y=FluxAvgRep, color=Site)) + geom_point() + geom_jitter(width
 coord_cartesian(ylim = c(0, 357))
sed_time
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_point()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_point()').
```
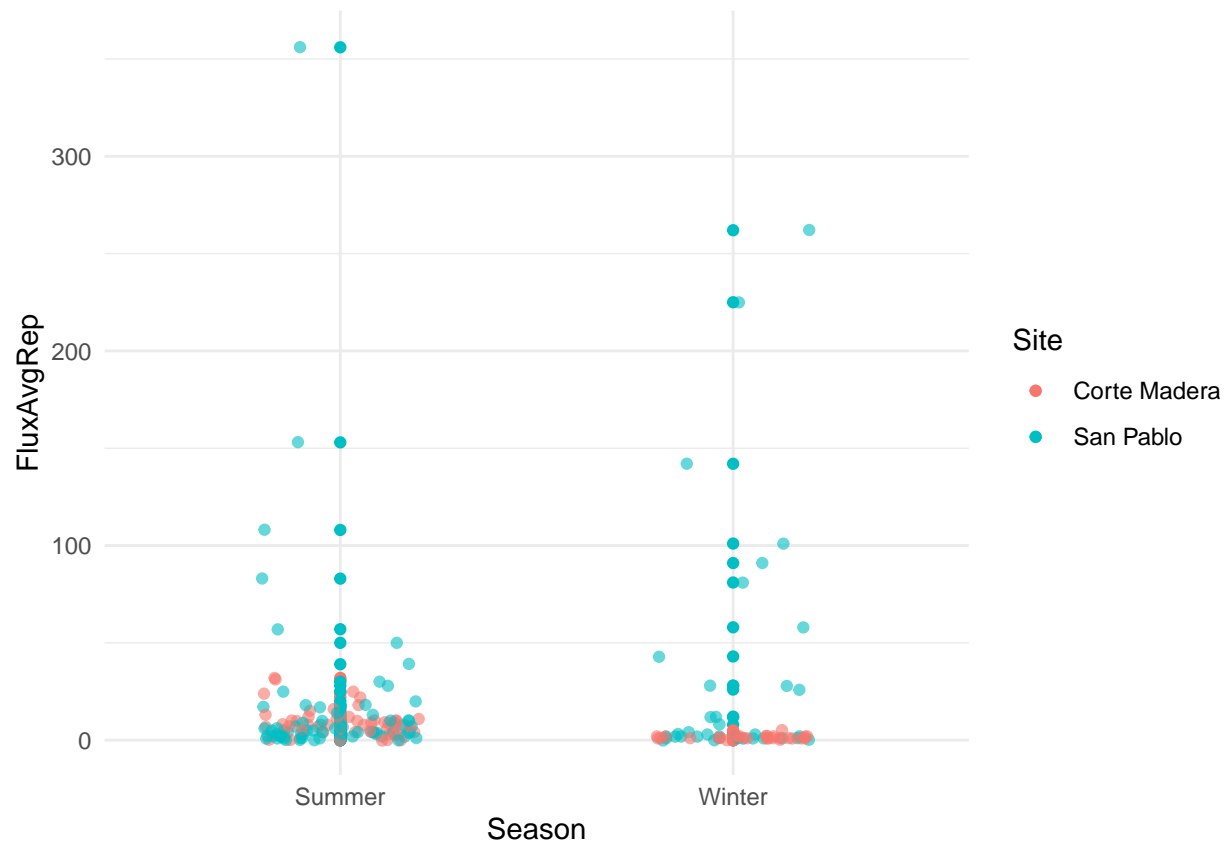
```
SAPA_height <- ggplot(veg, aes(x=Max.Ht,fill=Season)) + geom_boxplot()
SAPA_height
```
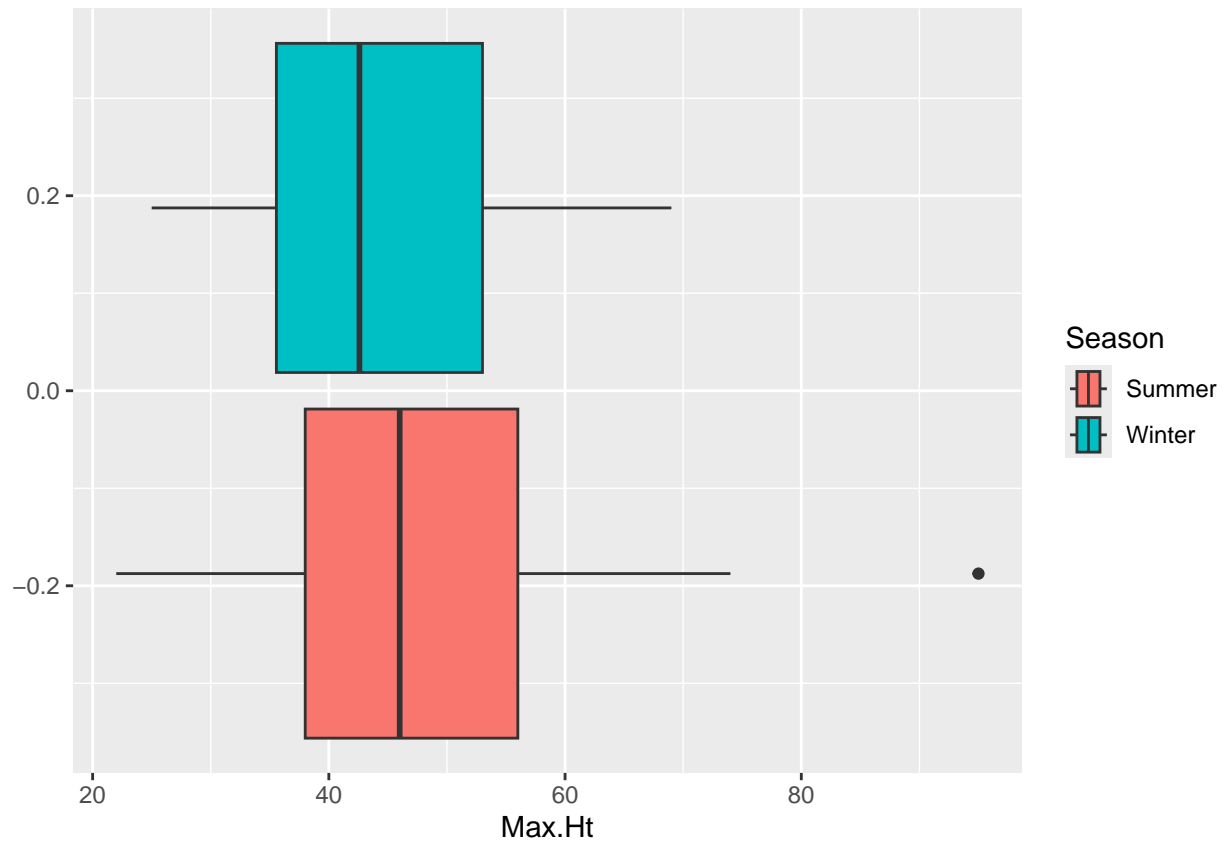
```
SAPA_cover <- ggplot(veg, aes(x=Cover, fill = Season)) + geom_boxplot()
SAPA_cover
```
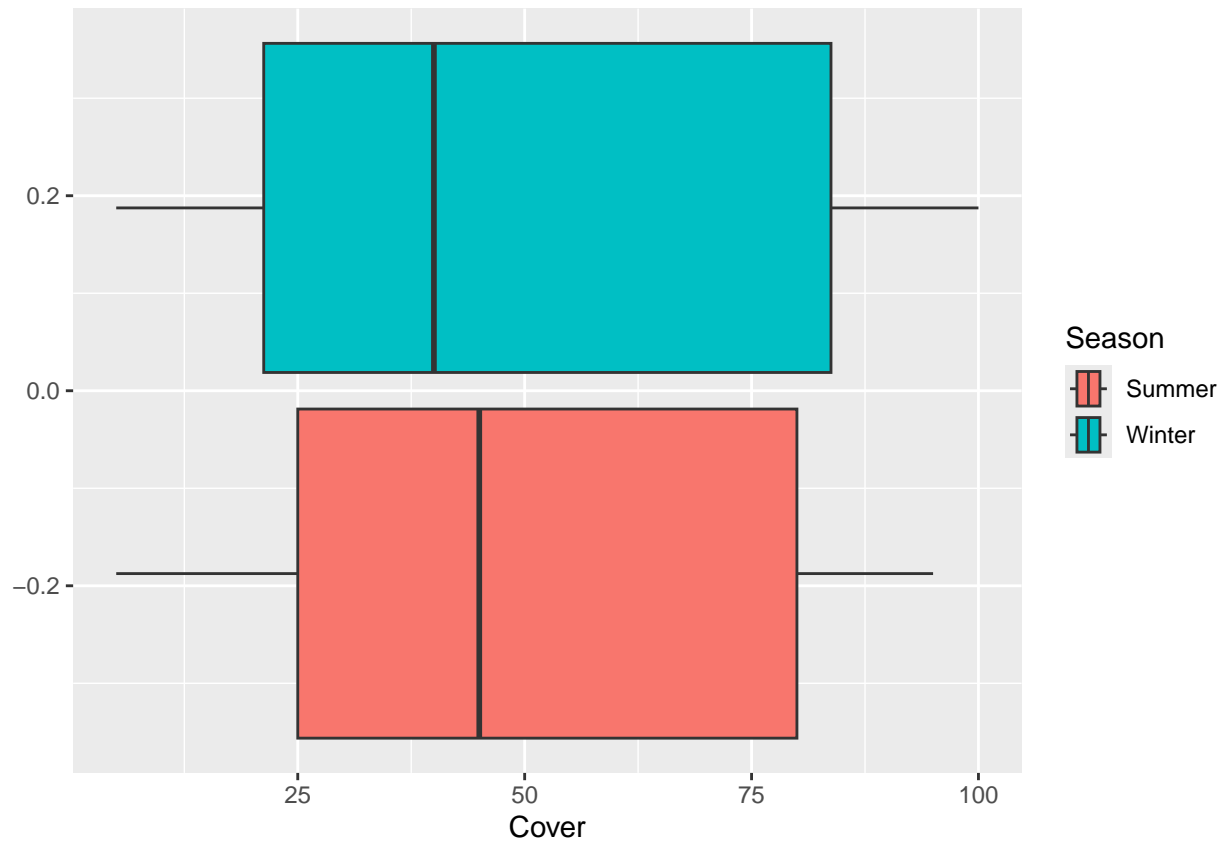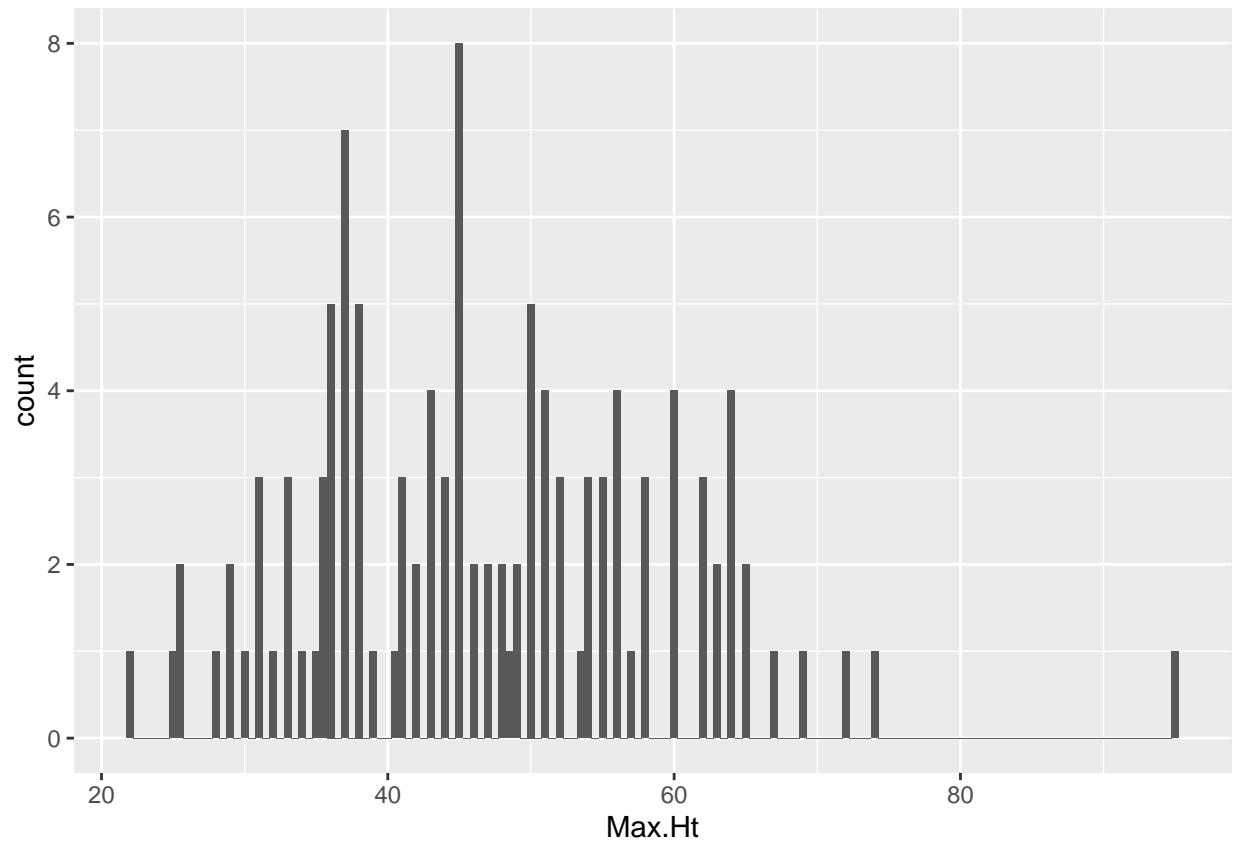
```
SAPA_Site_cover <- ggplot(veg, aes(x= Cover, fill = Site)) + geom_boxplot()
SAPA_Site_cover
```
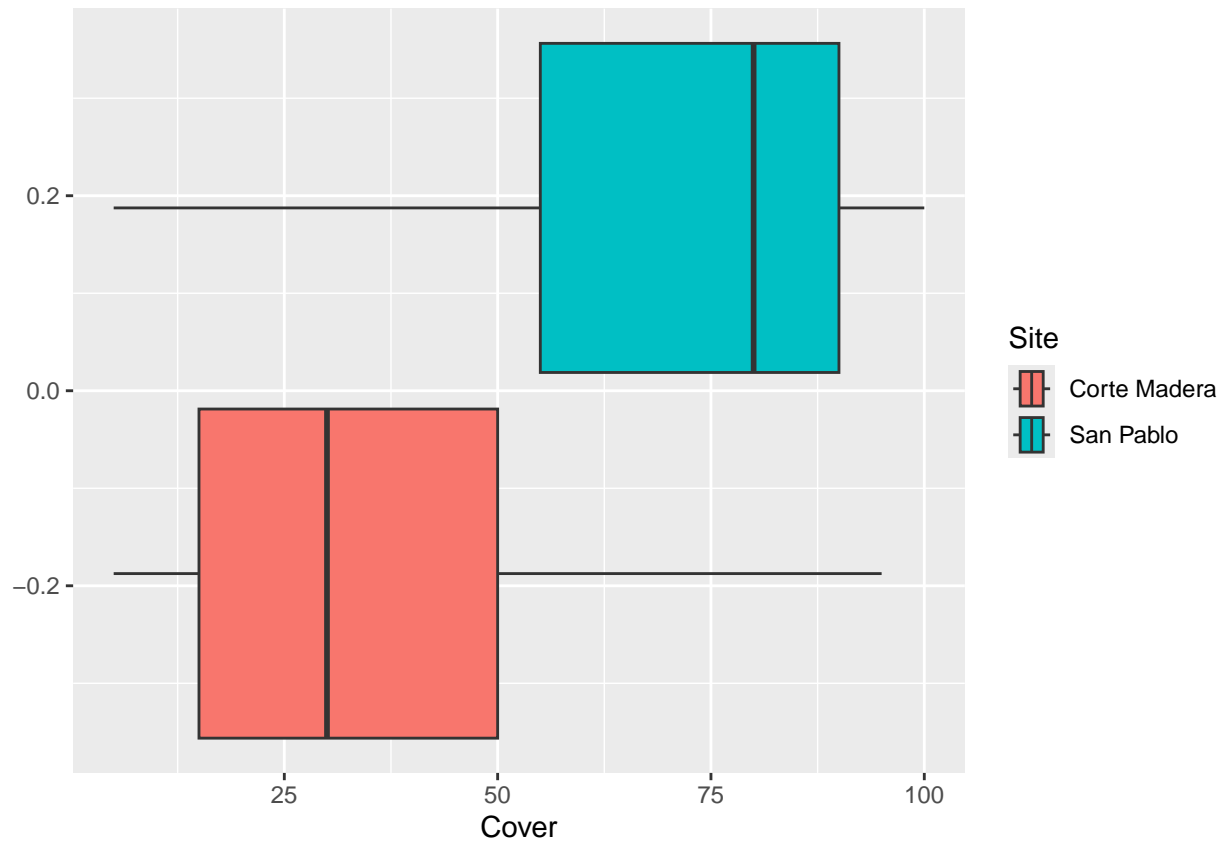
```
SAPA_Site_mxht <- ggplot(veg, aes(x = Max.Ht, fill = Site)) + geom_boxplot()
SAPA_Site_mxht
```
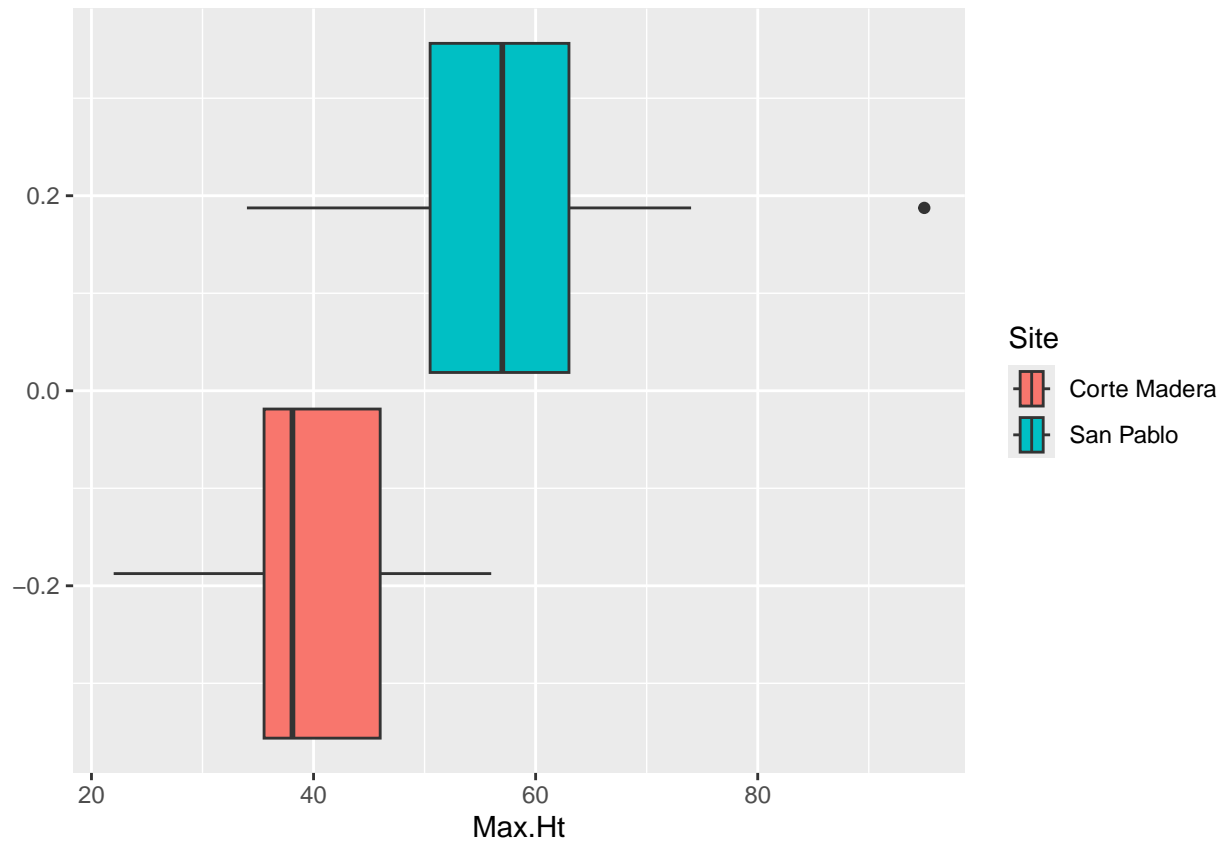
```
#check the data for normality, Max.Ht
SAPA_Maxht_Site_Normality <-ggplot(veg, aes(x=Max.Ht)) + geom_histogram(binwidth = 0.5)
SAPA_Maxht_Site_Normality
```

```
#check the data for normality, Avg.ht
SAPA_Avght_Site_Normality <-ggplot(veg, aes(x=Avg.Ht)) + geom_histogram(binwidth = 0.5)
SAPA_Avght_Site_Normality
```

```
#check the data for normality, Cover
SAPA_Cover_Site_Normality <-ggplot(veg, aes(x=Cover)) + geom_histogram(binwidth = 0.5)
SAPA_Cover_Site_Normality
```
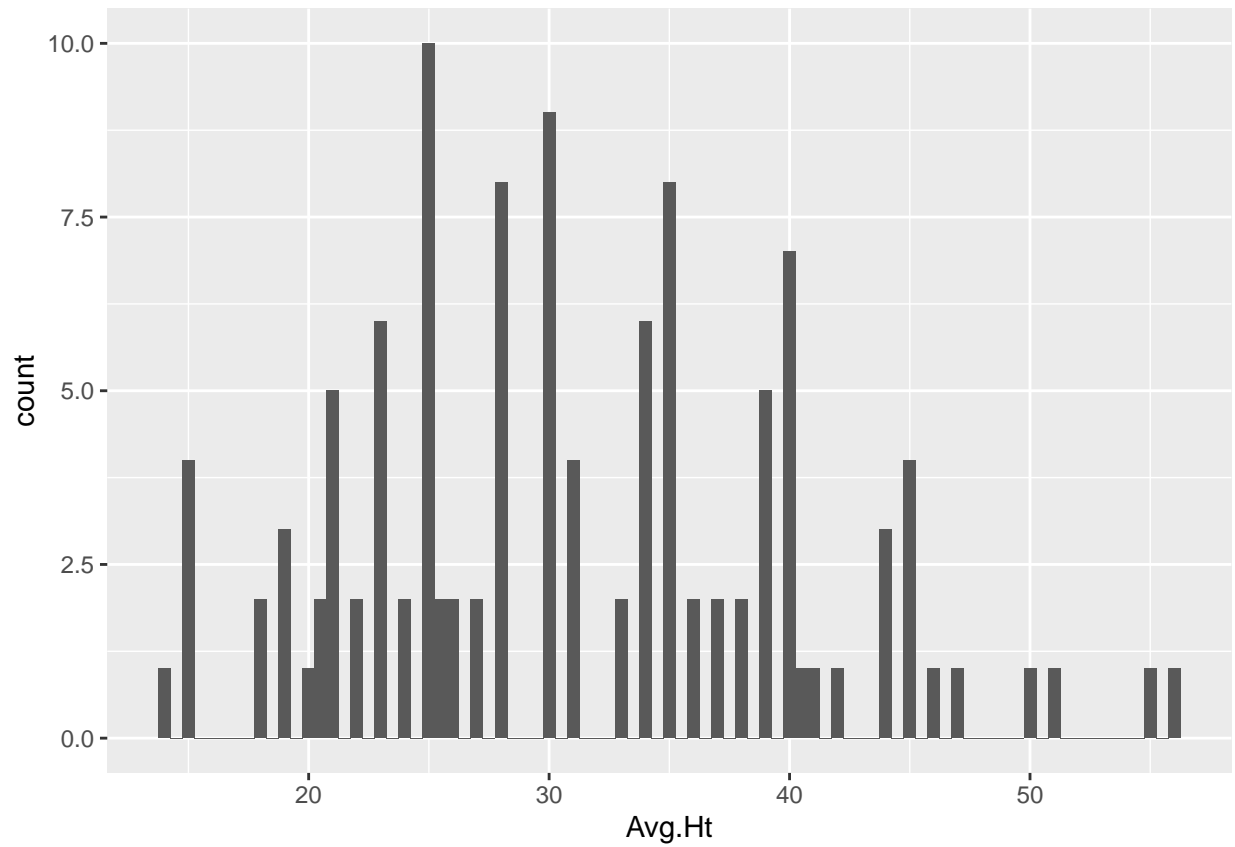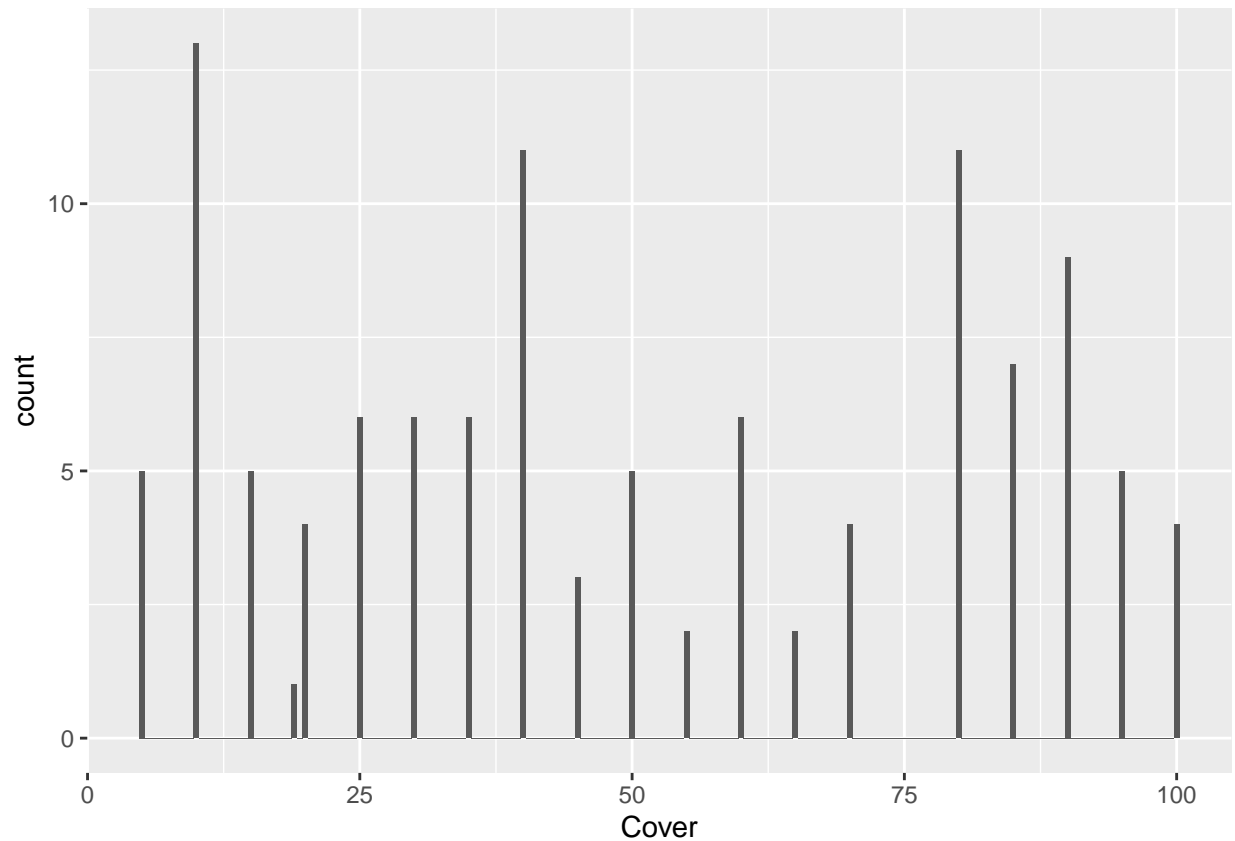
```
sed_SAPA <- merge(sed, veg, by = c("Site", "Season", "Distance", "Transect", "Year", "Month"), all = FAL

#separating the main dataframe into two site sets

CMsed_veg <- subset(sed_SAPA, Site == "Corte Madera")
#CMsed_veg

SPsed_veg <- subset(sed_SAPA, Site == "San Pablo")
#SPsed_veg
```
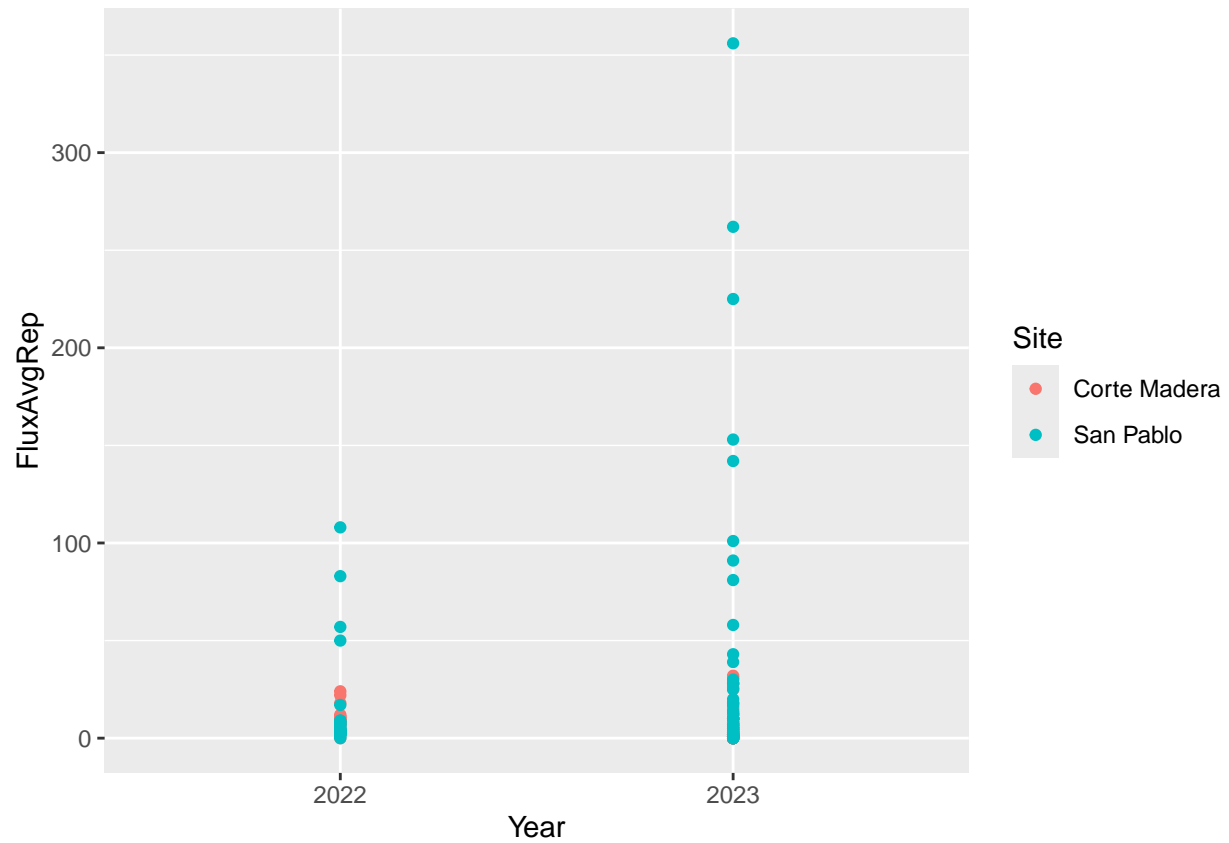
```
sed_year <- ggplot(sed, aes(x= Year, y= FluxAvgRep, color= Site)) +geom_point()
sed_year
```
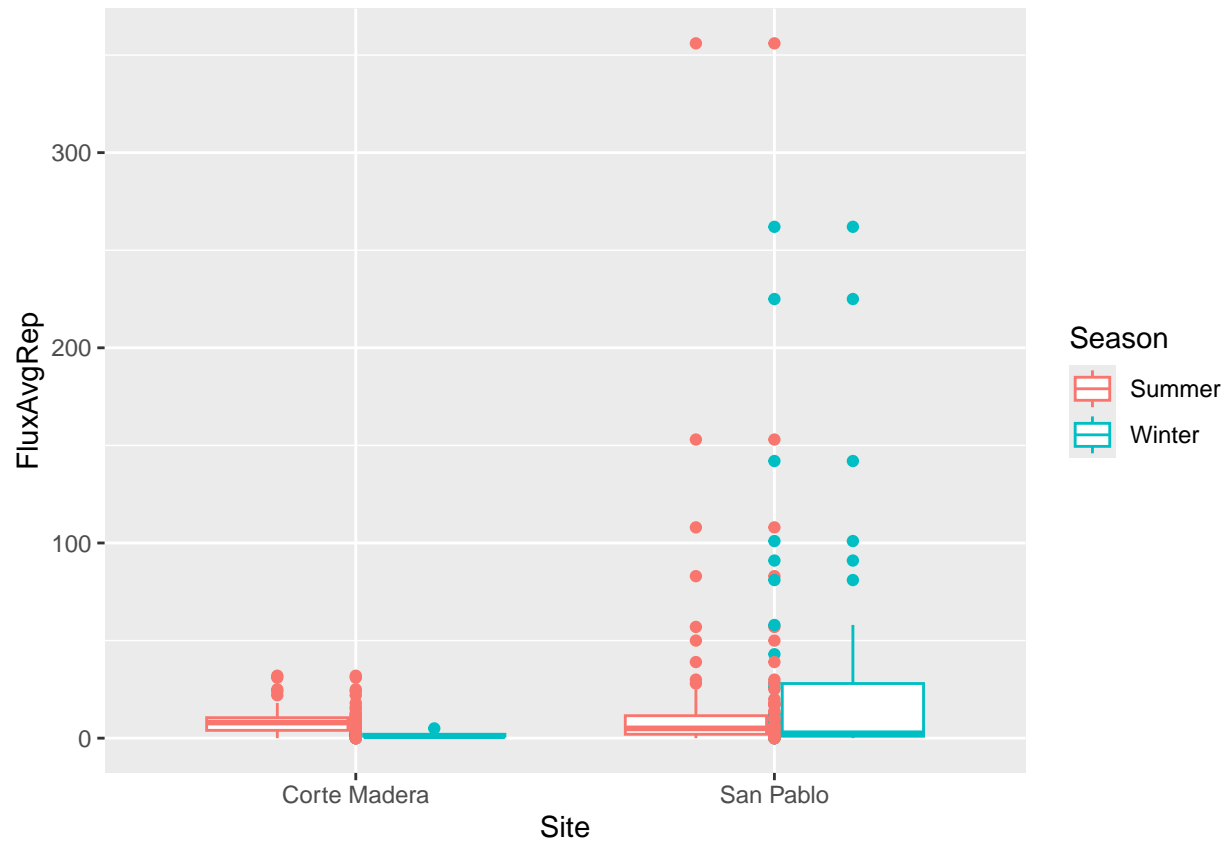
```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_point()').
```
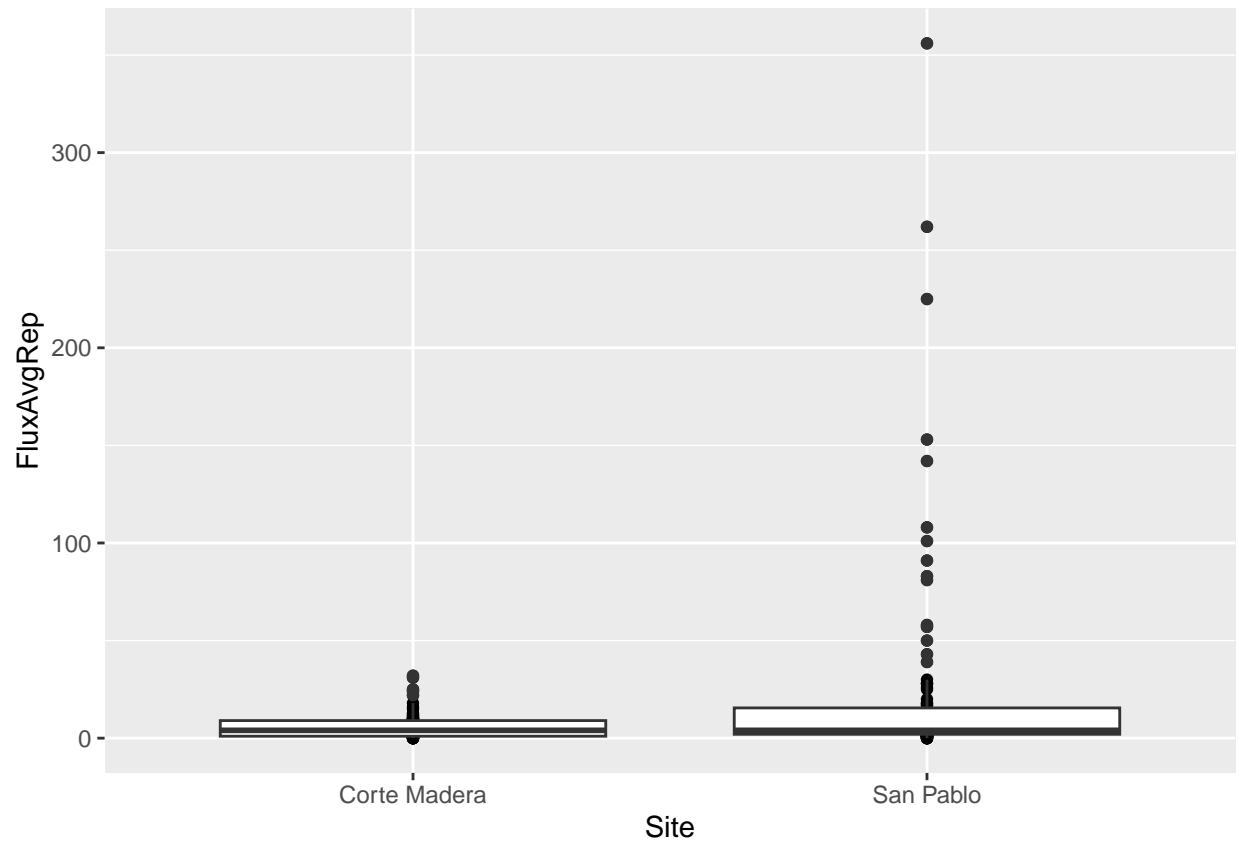
```
sed_Siteseason <- ggplot(sed, aes(x= Site, y=FluxAvgRep, color = Season)) +geom_point() + geom_boxplot(
sed_Siteseason
```

```
## Warning: Removed 1 row containing non-finite outside the scale range ('stat_boxplot()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_point()').
```
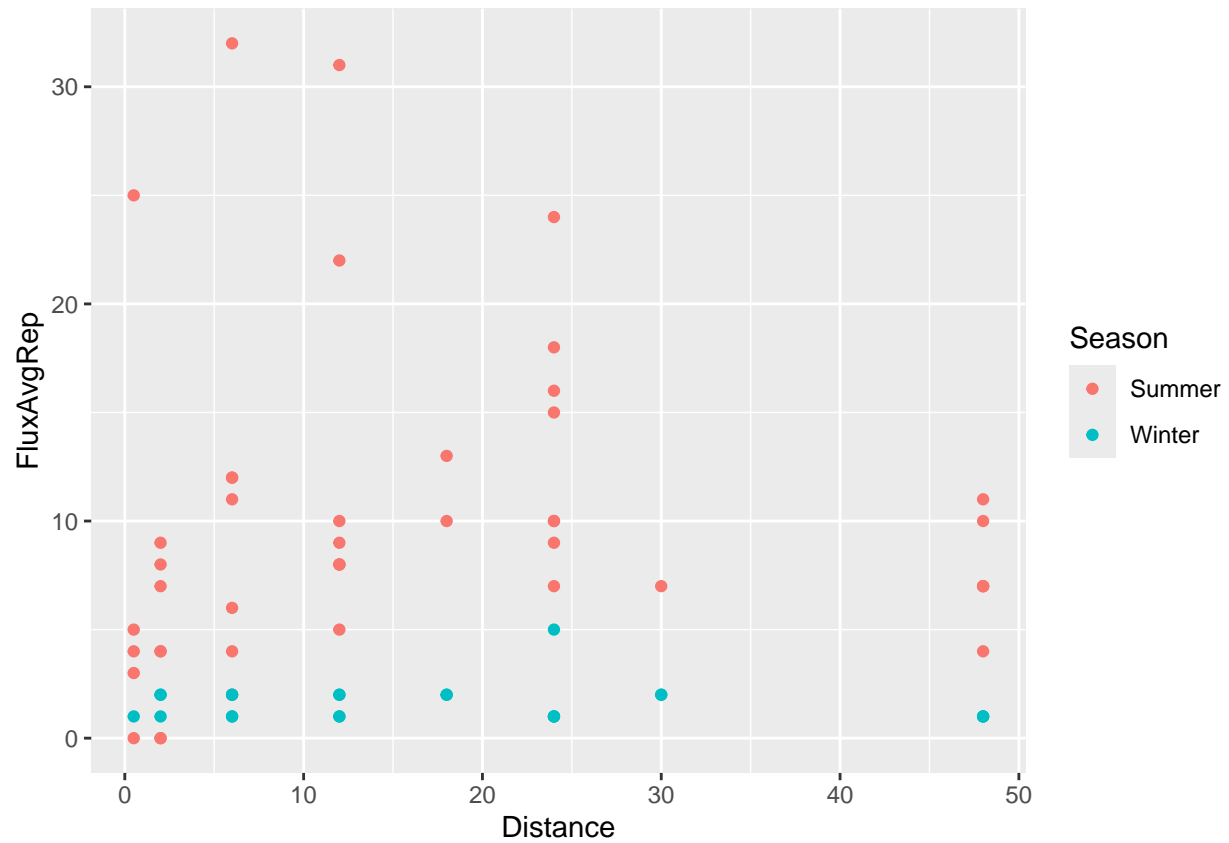
```
sed_site <- ggplot(sed, aes(x=Site, y=FluxAvgRep)) +geom_point() + geom_boxplot()
sed_site
```

```
## Warning: Removed 1 row containing non-finite outside the scale range ('stat_boxplot()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_point()').
```
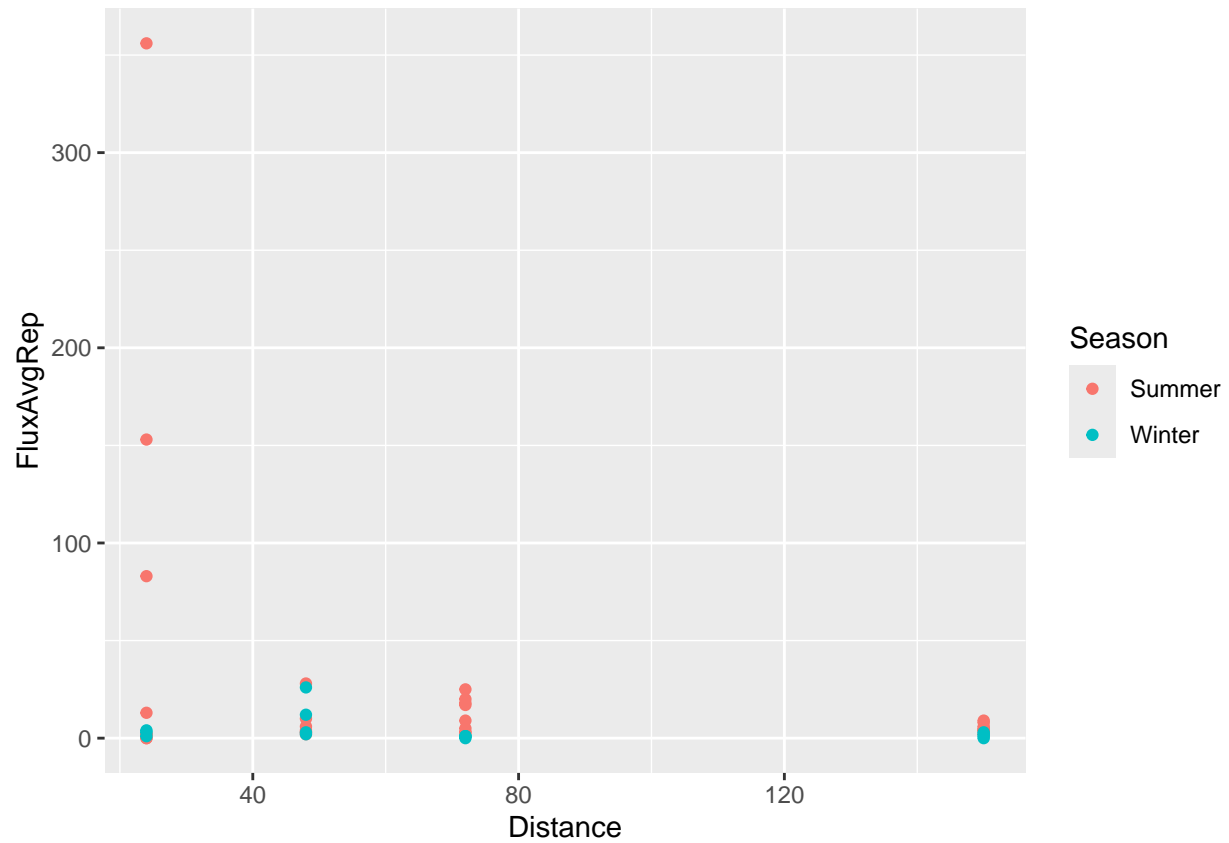
```
sedvegCM <- ggplot(CMsed_veg, aes(x = Distance, y = FluxAvgRep, color = Season)) + geom_point()
sedvegCM
```
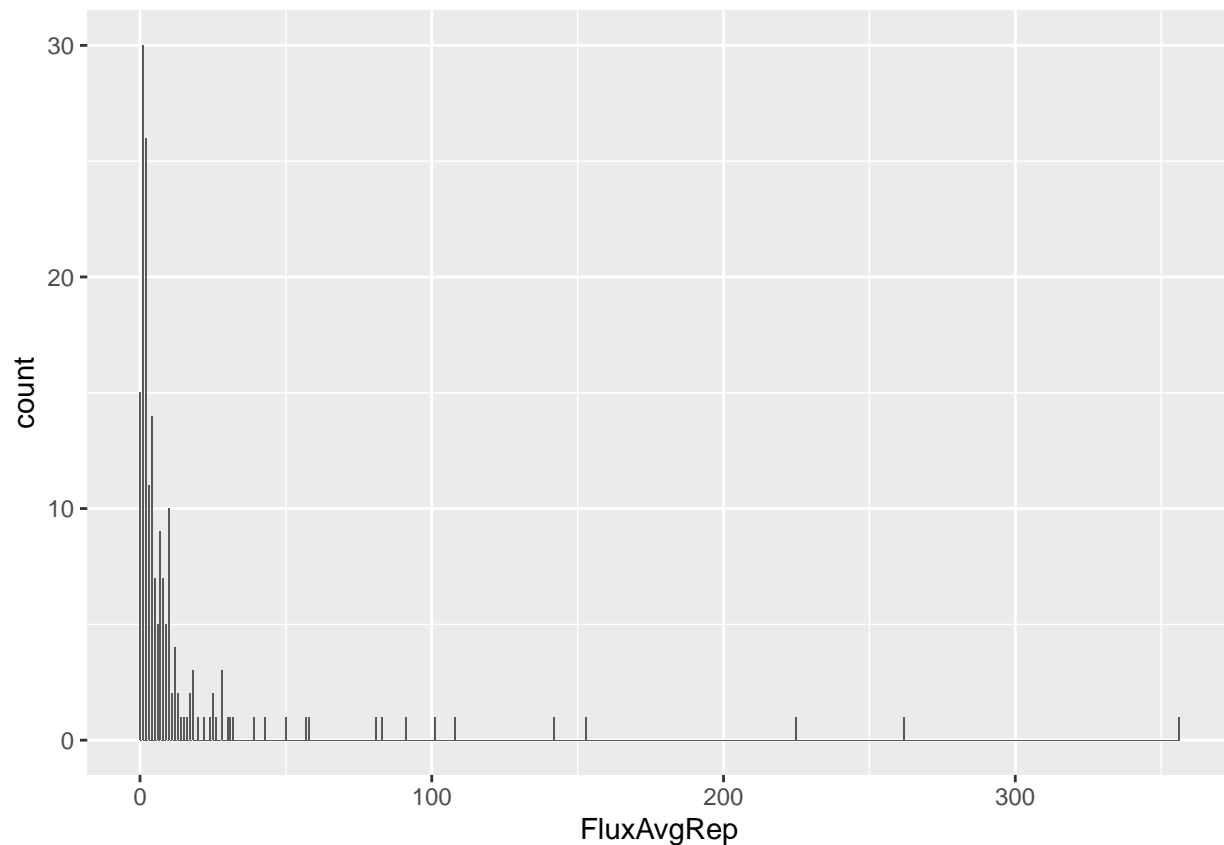
```
sedvegSP <- ggplot(SPsed_veg, aes(x = Distance, y = FluxAvgRep, color = Season)) + geom_point()
sedvegSP
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_point()').
```

```
sed_site_normality <- ggplot(sed, aes(x=FluxAvgRep)) +geom_histogram(binwidth = 0.5)
sed_site_normality
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## ('stat_bin()').
```

```
t.test(Max.Ht~Site,alternative ="two.sided", var.equal = TRUE, data=sed_SAPA)
```

```
##
##   Two Sample t-test
##
## data:  Max.Ht by Site
## t = -9.6305, df = 112, p-value = 2.355e-16
## alternative hypothesis: true difference in means between group Corte Madera and group San Pablo is ne
## 95 percent confidence interval:
##  -20.27192 -13.35377
## sample estimates:
## mean in group Corte Madera     mean in group San Pablo
##                    39.80418                    56.61702
```

```
t.test(FluxAvgRep~Site,alternative ="two.sided", var.equal = TRUE, data=sed_SAPA)
```

```
##
##   Two Sample t-test
##
## data:  FluxAvgRep by Site
## t = -1.6895, df = 111, p-value = 0.09394
## alternative hypothesis: true difference in means between group Corte Madera and group San Pablo is ne
## 95 percent confidence interval:
##  -25.679157    2.043206
```

```
## sample estimates:
## mean in group Corte Madera    mean in group San Pablo
##                  7.029851                  18.847826
```

```
sedvegglm <-glm(FluxAvgRep~Max.Ht, family = "poisson", data=sed_SAPA)
sedvegglm
```

```
##
## Call:  glm(formula = FluxAvgRep ~ Max.Ht, family = "poisson", data = sed_SAPA)
##
## Coefficients:
## (Intercept)        Max.Ht
##     2.634206     -0.003517
##
## Degrees of Freedom: 112 Total (i.e. Null);  111 Residual
##    (1 observation deleted due to missingness)
## Null Deviance:        3415
## Residual Deviance: 3412  AIC: 3788
```
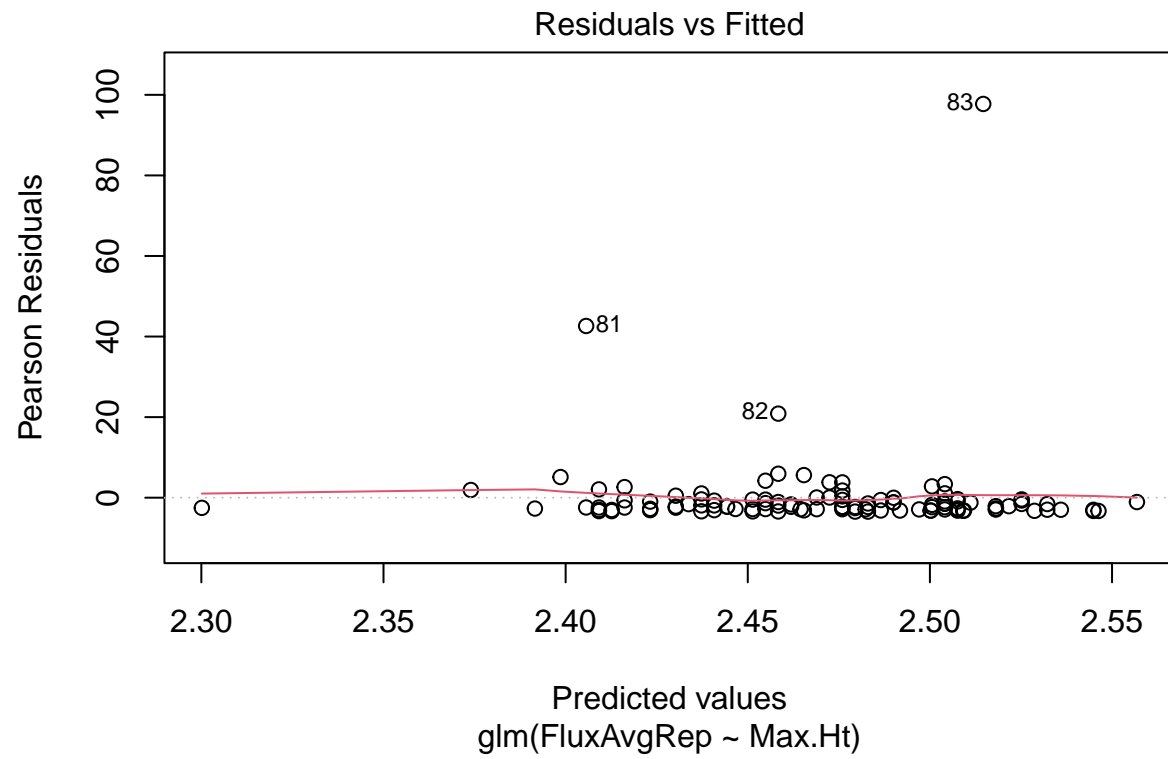
```
sedvegglm_AIC_text <- paste("AIC: 3788")
sedvegglm_AIC_text
```
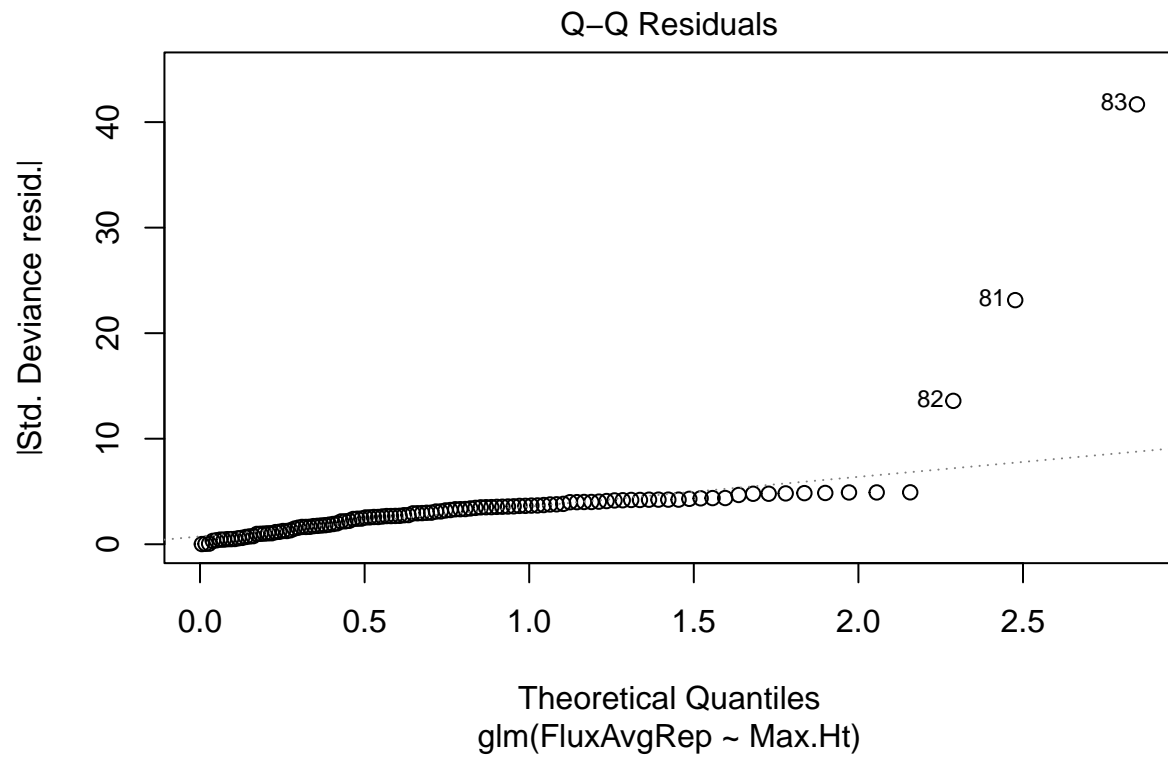
```
## [1] "AIC: 3788"
```

```
sedvegglm_coefficient_Max.Ht <- paste("Coefficient for Max Height: -0.003517 ")
sedvegglm_coefficient_Max.Ht
```

```
## [1] "Coefficient for Max Height: -0.003517 "
```

```
plot(sedvegglm)
```

## Residuals vs Fitted



Pearson Residuals

83

81

82

Predicted values
glm(FluxAvgRep ~ Max.Ht)

Q–Q Residuals

Theoretical Quantiles
glm(FluxAvgRep ~ Max.Ht)

Scale–Location

√|Std. Pearson resid.|

Predicted values
glm(FluxAvgRep ~ Max.Ht)

## Residuals vs Leverage



Leverage
glm(FluxAvgRep ~ Max.Ht)
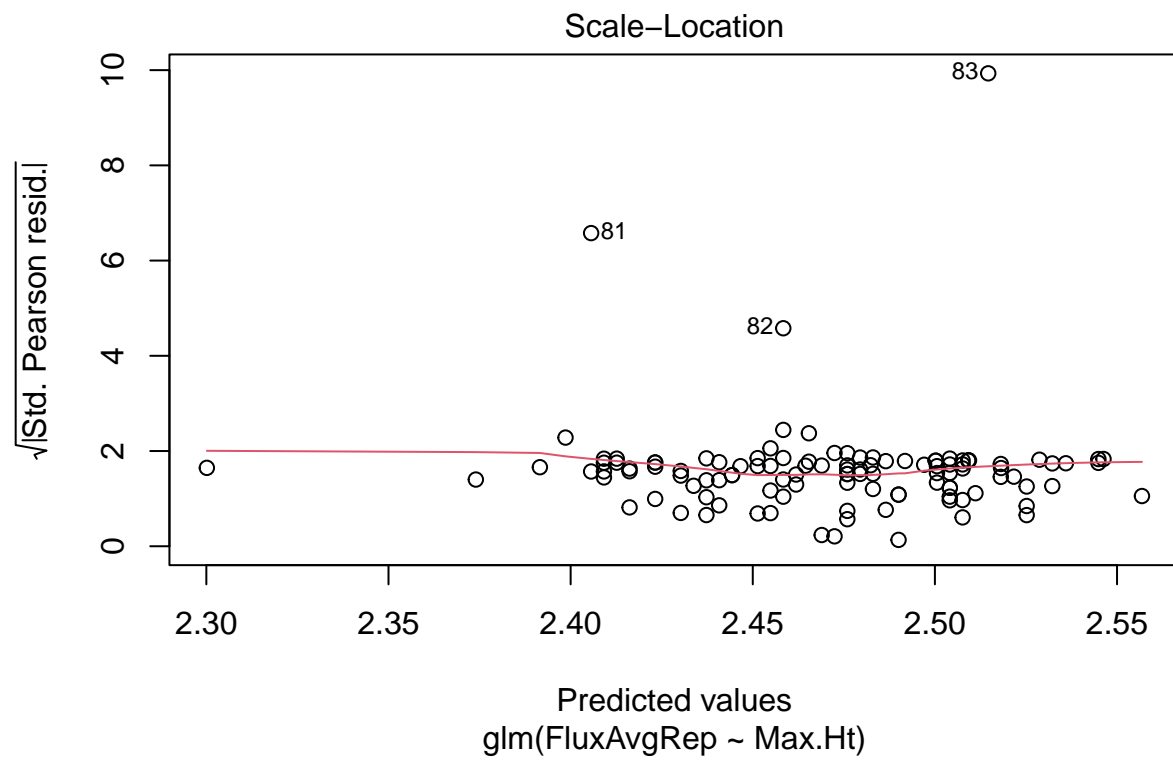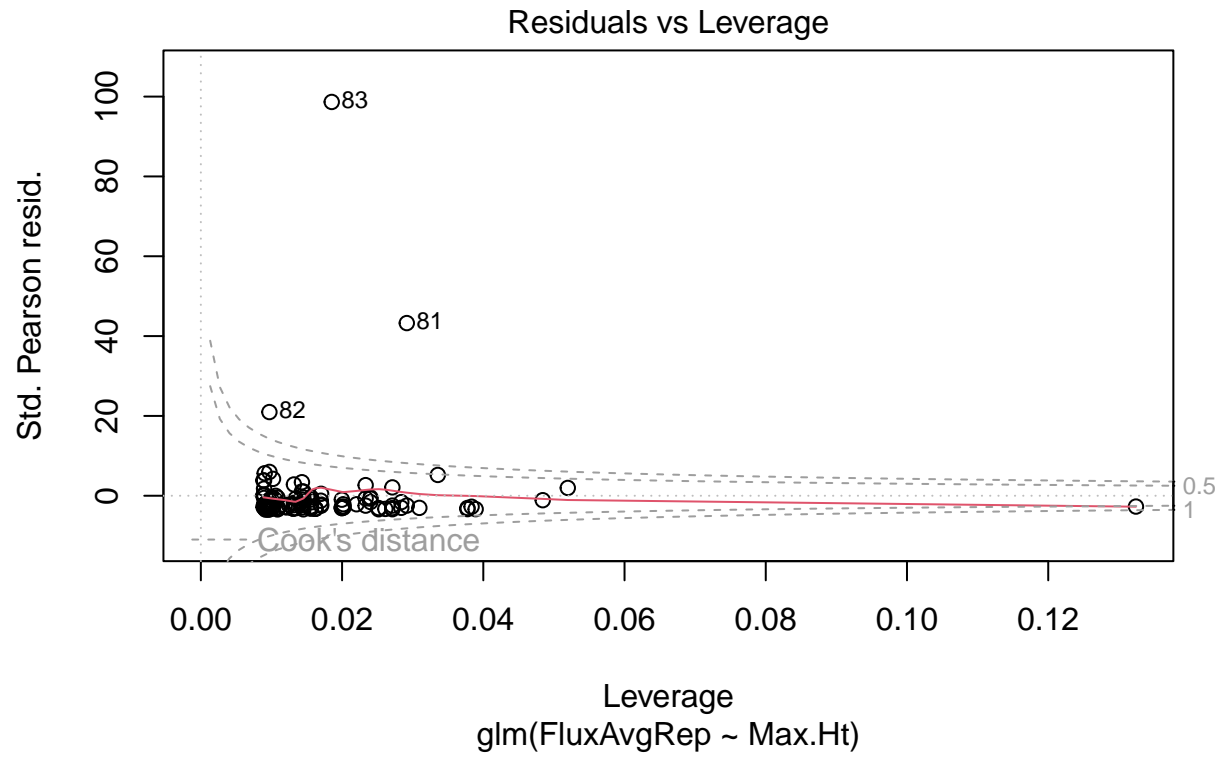
```
library(glmmTMB)
```

```
## Warning: package 'glmmTMB' was built under R version 4.4.3
```

```
sed_glmm = glmmTMB(FluxAvgRep~Max.Ht + Distance + Avg.Ht + (1|Season) + (1|Site) +(1|Year), family = po:
summary(sed_glmm)
```

```
##  Family: poisson  ( log )
## Formula:
## FluxAvgRep ~ Max.Ht + Distance + Avg.Ht + (1 | Season) + (1 |
##     Site) + (1 | Year)
## Data: sed_SAPA
##
##      AIC      BIC    logLik -2*log(L)  df.resid
##    1621.2   1640.3    -803.6    1607.2       106
##
## Random effects:
##
## Conditional model:
##  Groups Name        Variance Std.Dev.
##  Season (Intercept) 1.4483   1.2034
##  Site   (Intercept) 1.4312   1.1963
##  Year   (Intercept) 0.4267   0.6532
## Number of obs: 113, groups:  Season, 2; Site, 2; Year, 2
```

```
##
## Conditional model:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.488282   1.292109   3.474 0.000514 ***
## Max.Ht      -0.114653   0.006127 -18.714  < 2e-16 ***
## Distance    -0.018136   0.001113 -16.298  < 2e-16 ***
## Avg.Ht       0.094242   0.007535  12.507  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
AIC(sed_glmm)
```

```
## [1] 1621.219
```

```r
sed_glmmSP = glmmTMB(FluxAvgRep~Max.Ht + Distance + (1|Season) + (1|Year), family = poisson, data=SPsed_
summary(sed_glmmSP)
```

```
##  Family: poisson  ( log )
## Formula:          FluxAvgRep ~ Max.Ht + Distance + (1 | Season) + (1 | Year)
## Data: SPsed_veg
##
##      AIC      BIC   logLik -2*log(L)  df.resid
##   1152.6   1161.7   -571.3    1142.6        41
##
## Random effects:
##
## Conditional model:
##  Groups Name        Variance Std.Dev.
##  Season (Intercept) 1.7208   1.3118
##  Year   (Intercept) 0.9746   0.9872
## Number of obs: 46, groups:  Season, 2; Year, 2
##
## Conditional model:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  5.890846   1.172604   5.024 5.07e-07 ***
## Max.Ht      -0.054959   0.003069 -17.907  < 2e-16 ***
## Distance    -0.022194   0.001291 -17.198  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
sed_glmmCM = glmmTMB(FluxAvgRep~Max.Ht + Distance + (1|Season) + (1|Year), family = poisson, data=CMsed_
summary(sed_glmmCM)
```

```
##  Family: poisson  ( log )
## Formula:          FluxAvgRep ~ Max.Ht + Distance + (1 | Season) + (1 | Year)
## Data: CMsed_veg
##
##      AIC      BIC   logLik -2*log(L)  df.resid
##    470.7    481.7   -230.4     460.7        62
##
## Random effects:
##
```

```
## Conditional model:
##  Groups Name          Variance  Std.Dev.
##  Season (Intercept) 7.687e-01 8.768e-01
##  Year   (Intercept) 6.767e-10 2.601e-05
## Number of obs: 67, groups:  Season, 2; Year, 2
##
## Conditional model:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.811453   0.679949   1.193    0.233
## Max.Ht      0.012982   0.006290   2.064    0.039 *
## Distance    0.003845   0.003062   1.256    0.209
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
sed1_cor <- cor_test(FluxAvgRep, Max.Ht, data=sed_SAPA)

#correlation coefficient
r <- sed1_cor$cor


r
```

```
##     cor
## -0.014
```

```r
# plotting the interaction
p5 <- ggplot(sed_SAPA,aes(Max.Ht,FluxAvgRep,group=Season, color=Site)) + geom_point(size=2,alpha=.5,pos
  geom_smooth(method = "lm", se = FALSE) +
  ylab("Sediment Flux") +
  xlab("Maximum Plant Height") +
  theme_classic(14)
p5
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## ('stat_smooth()').
```

```
## Warning: The following aesthetics were dropped during statistical transformation:
## colour.
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a 'group' aesthetic or to convert a numerical
##   variable into a factor?
```

```
## Warning: 'position_dodge()' requires non-overlapping x intervals.
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_point()').
```