

# Supplementary File 2

20 April 2024

This file documents our reanalysis of the dataset used to examine biodiversity and conflict relationships in the Southern Philippines presented in this paper <https://doi.org/10.1038/s44185-024-00044-8>

## Load required packages

```
#make sure to install all packages including required dependencies prior to use
library(dplyr)
library(tidyverse)
library(ggplot2)
library(ggpubr)
library(rgdal)
library(raster)
library(lme4)
library(MASS)
library(piecewiseSEM)
```

## Prepare datasets

```
#biodiversity data was obtained from here https://doi.org/10.15468/rtedgk
#and conflict data was from here https://data.humdata.org/dataset/ucdp-data-for-philippines?
#note that only a subset of this dataset (i.e., those from Mindanao and adjacent islands) was used

#biodiversity data
bio <- read.csv("mobios_tab.csv", header = TRUE, sep = ",")
#subset data
bio.sub <- subset(bio, select = c(12, 23, 20, 15, 16))
bio.sub <- bio.sub %>% filter(county!="") #remove rows with no county information
bio.sub <- bio.sub %>% filter(class!="Malacostraca") #exclude this taxon
bio.sub <- bio.sub %>% filter(class!="Bivalvia") #exclude this taxon
bio.sub <- bio.sub %>% filter(class!="Gastropoda")#exclude this taxon
#check province name
#there are rows where the name of the province is uncertain as indicated by "/"
##"Zamboanga-fdel Norte" was changed to "Zamboanga del Norte" prior to import of data
unique(bio.sub$county)

## [1] "Lanao del Norte"
## [2] "South Cotabato"
## [3] "North Cotabato"
## [4] "Camiguin Island"
## [5] "Maguindanao"
## [6] "Agusan del Sur"
## [7] "Bukidnon"
## [8] "Davao Oriental"
## [9] "Davaao"
## [10] "Misamis Occidental"
## [11] "Surigao del Sur"
## [12] "Davaao del sur"
```

```

## [13] "Lanao del Sur"
## [14] "Sarangani"
## [15] "Misamis Oriental"
## [16] "Surigao del Norte"
## [17] "Davao de Oro"
## [18] "Davao del Norte"
## [19] "Zamboanga del Sur"
## [20] "Surigao del Norte/Agusan del Sur"
## [21] "Dinagat Island"
## [22] "Basilan"
## [23] "Tawi-Tawi"
## [24] "Sulu"
## [25] "Sultan Kudarat"
## [26] "Agusan del Norte"
## [27] "Cagayan de Oro"
## [28] "Bukidnon/Camiguin"
## [29] "Bukidnon/Misamis Oriental/Iligan"
## [30] "Zamboanga del Norte"
## [31] "Zamboanga Sibugay"
## [32] "Misamis Occidental/Misamis Oriental/Camiguin/Bukidnon"
## [33] "Zamboanga City"
## [34] "General Santos"
## [35] "North Cotabato/Davao City"

#remove uncertain province names
bio.sub <- bio.sub %>% filter(!grep('/', county))

```

```
head(bio.sub)
```

```

##           county   class      scientificName decimalLatitude
## 1 Lanao del Norte     Aves      Buceros hydrocorax    8.2122
## 2 Lanao del Norte  Mammalia    Eonycteris spelaea    8.3218
## 3 Lanao del Norte  Mammalia    Hipposideros diadema    8.3218
## 4 Lanao del Norte  Mammalia    Pipistrellus javanicus    8.3218
## 5 Lanao del Norte  Mammalia    Rhinolophus inops    8.3218
## 6 Lanao del Norte  Mammalia Rousettus amplexicaudatus    8.3218
##   decimalLongitude
## 1          124.3835
## 2          124.2940
## 3          124.2940
## 4          124.2940
## 5          124.2940
## 6          124.2940

```

```

#conflict data
con <- read.csv("conflict_tab_89-21.csv", header = TRUE, sep = ",")
con.sub <- subset(con, select = c(1,3,15,18,30,32,33))

```

```
head(con.sub)
```

```

##      id year         side_a side_b      PROVINCE latitude
## 1 124029 1989 Government of Philippines CPP      Cotabato 6.951400
## 2 125199 1989 Government of Philippines CPP      Maguindanao 7.270000
## 3 123871 1989 Government of Philippines CPP Surigao del Sur 8.666670
## 4 122819 1989 Government of Philippines CPP Agusan del Sur 8.500000
## 5 123283 1989 Government of Philippines CPP Lanao del Norte 8.000000

```

```

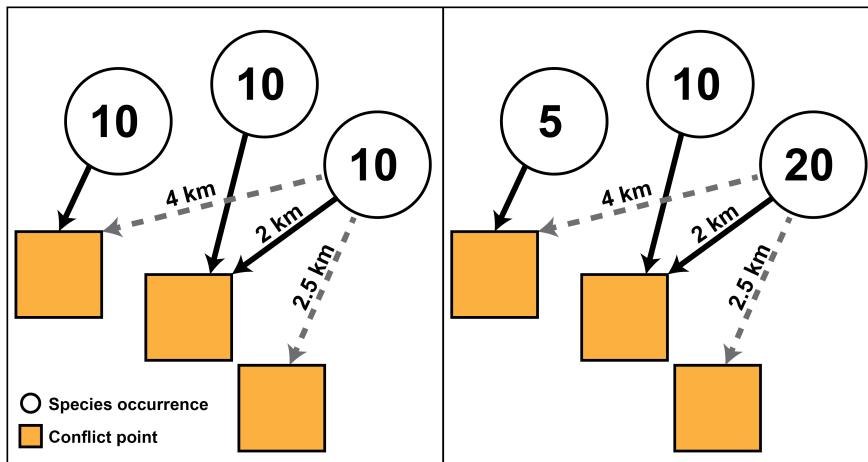
## 6 123344 1989 Government of Philippines      CPP  Davao Oriental 7.081413
##   longitude
## 1 125.0992
## 2 124.3106
## 3 126.0000
## 4 125.8333
## 5 124.0000
## 6 126.1299
write.csv(con.sub, "con.sub.csv")

```

Count recorded species per taxon in each point and within province

*#the methods of how species counts were done in each site was not provided in the paper  
#according to the paper analysis was done at the provincial level  
#there could be two different ways how the species counts were done (see figure below)  
#this part is extremely important which the authors failed to discuss  
#the left panel of the figure shows that species counts were aggregated at the provincial level  
#each point (within the province) receives the same value of species count (or species richness)  
#the right panel shows that each point has a unique number of species count  
#note that these values are crucial for other downstream analyses  
#particularly when relationships of species counts  
#and distance to the nearest conflict sites are considered  
#we generated the values for these two scenarios below  
#and test whether any of them would match the results presented in the paper*

```
knitr:::include_graphics("bioconflict.png")
```



```
#count unique records (i.e., species richness) per point
bio.data.sum <- bio.sub %>%
  group_by(county, class, decimalLatitude, decimalLongitude) %>%
  mutate(NumbSp = n_distinct(scientificName))
```

```
head(bio.data.sum)
```

```

## # A tibble: 6 x 6
## # Groups:   county, class, decimalLatitude, decimalLongitude [2]
##   county       class   scientificName decimalLatitude decimalLongitude NumbSp
##   <chr>        <chr>    <chr>           <dbl>            <dbl>    <int>
## 1 Lanao del Norte Aves     Buceros hydro~      8.21            124.     1
## 2 Lanao del Norte Mammal~ Eonycteris sp~      8.32            124.     5

```

```

## 3 Lanao del Norte Mammal~ Hipposideros ~          8.32      124.      5
## 4 Lanao del Norte Mammal~ Pipistrellus ~         8.32      124.      5
## 5 Lanao del Norte Mammal~ Rhinolophus i~         8.32      124.      5
## 6 Lanao del Norte Mammal~ Rousettus amp~        8.32      124.      5
#export data
write.csv(bio.data.sum, "bio.data.sum.csv")

#count species records per province (this lumps all the data,
#thus each point will have a single species richness value)
prov.bio.data.sum <- bio.sub %>%
  group_by(county, class) %>%
  mutate(NumbSp = n_distinct(scientificName))
head(prov.bio.data.sum)

## # A tibble: 6 x 6
## # Groups:   county, class [2]
##   county       class scientificName decimalLatitude decimalLongitude NumbSp
##   <chr>        <chr>    <chr>           <dbl>            <dbl>    <int>
## 1 Lanao del Norte Aves   Buceros hydro~      8.21             124.     1
## 2 Lanao del Norte Mammal~ Eonycteris sp~     8.32             124.     5
## 3 Lanao del Norte Mammal~ Hipposideros ~     8.32             124.     5
## 4 Lanao del Norte Mammal~ Pipistrellus ~     8.32             124.     5
## 5 Lanao del Norte Mammal~ Rhinolophus i~     8.32             124.     5
## 6 Lanao del Norte Mammal~ Rousettus amp~     8.32             124.     5
#export data
write.csv(prov.bio.data.sum, "prov.data.sum.csv")

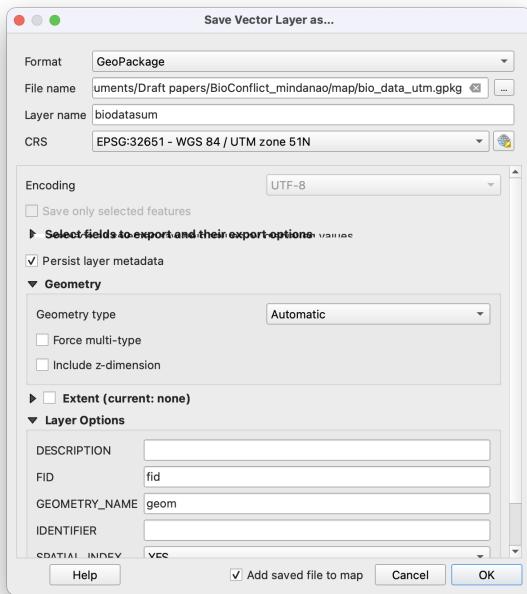
```

Calculate the distance of species record to the nearest conflict site/s using QGIS

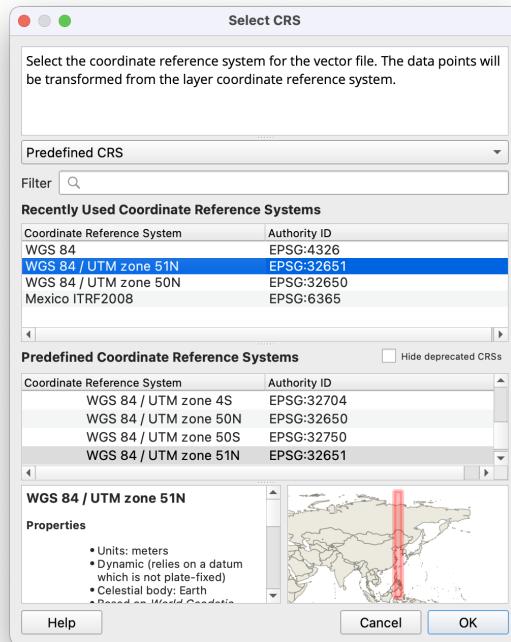
```

#in QGIS, import the "bio.data.sum" and "con.sub" from R as delimited text files.
#export these files as GeoPackage or Shapefile to convert the data to meters prior to calculation
#use the CRS UTM Zone 51N
knitr:::include_graphics("qgis_geo.png")

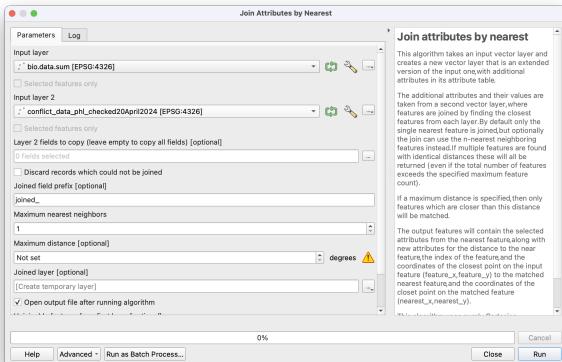
```



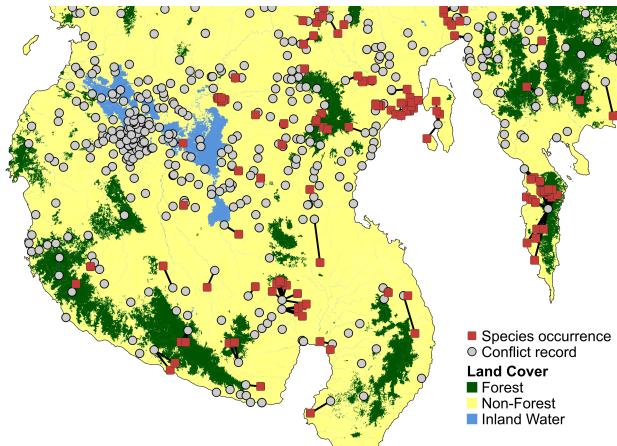
```
knitr:::include_graphics("qgis_utm.png")
```



```
#import back the newly saved data to QGIS
#get the distance of nearest conflict areas using the "Joint attributes by nearest" plugin
#QGIS > Processing Toolbox > Join attributes by nearest
knitr:::include_graphics("qgis_join.png")
```



```
#distance can be calculated as well without joining the attributes using the "Distance to nearest hub"
#after calculating the distance, a new layer will be created.
#the new layer contains the calculated distance in the attribute table
#this also includes the x and y coordinates of the nearest site/point
#then export this new layer as a csv file so we can use the data in R for other analysis
#repeat the entire process for provincial data (i.e., prov.bio.data.sum)
#figure below shows the nearest conflict point to species record (black line)
#notice the remainder of conflict points are not included
knitr:::include_graphics("hub.png")
```



### Calculate the average distance

```
#species richness per point in each province
#read the file with calculated distance from QGIS
bio.con.dist <- read.csv("bio.con.dist.csv", header = TRUE, sep = ",")
#get average distance by province for each taxon
mean.dist <- bio.con.dist %>%
  group_by(county, class, NumbSp, decimalLatitude, decimalLongitude) %>%
  summarize(meanDistance = mean(distance, na.rm = TRUE))

## `summarise()` has grouped output by 'county', 'class', 'NumbSp',
## 'decimalLatitude'. You can override using the `groups` argument.

head(mean.dist)

## # A tibble: 6 x 6
## # Groups:   county, class, NumbSp, decimalLatitude [6]
##   county      class    NumbSp decimalLatitude decimalLongitude meanDistance
##   <fct>       <fct>    <dbl>        <dbl>          <dbl>
```

```

##   <chr>      <chr>     <int>     <dbl>     <dbl>     <dbl>
## 1 Agusan del Norte Amphibia    5       9.00     126.     5108.
## 2 Agusan del Norte Amphibia   20      8.97     126.     1594.
## 3 Agusan del Norte Amphibia   20      9.05     126.     8659.
## 4 Agusan del Norte Amphibia   20      9.05     126.     8052.
## 5 Agusan del Norte Amphibia   20      9.10     126.     2949.
## 6 Agusan del Norte Amphibia   20      9.15     126.     5850.

write.csv(mean.dist, "mean.distance.csv")

#species richness lumped per province
#read the file with calculated distance in QGIS
prov.bio.con.dist <- read.csv("prov.bio.con.dist.csv", header = TRUE, sep = ",")
#get average distance by province for each taxon
prov.mean.dist <- prov.bio.con.dist %>%
  group_by(county, class, NumbSp, decimalLatitude, decimalLongitude) %>%
  summarize(meanDistance = mean(distance, na.rm = TRUE))

## `summarise()` has grouped output by 'county', 'class', 'NumbSp',
## 'decimalLatitude'. You can override using the ` `.groups` argument.
write.csv(prov.mean.dist, "prov.mean.distance.csv")

#get average distance by province considering long/lat data (i.e., lumped average)
prov.lmp.dist <- prov.bio.con.dist %>%
  group_by(county, class, NumbSp) %>%
  summarize(lmpMeanDistance = mean(distance, na.rm = TRUE))

## `summarise()` has grouped output by 'county', 'class'. You can override using
## the ` `.groups` argument.
write.csv(prov.lmp.dist, "prov.lmp.distance.csv")

```

### Plot species record vs average distance

```

key.lab <- unique(mean.dist$class)
plot.color <- c("#7b3dcc", "#cc3d3d", "#000000", "#d970cb", "#8c994d", "#7f404a", "#67E3FC")
point.shape <- c(0,1,2,3,4,5,6,7,8,9,10)

#plot data with species richness per point
p1 <- ggplot(mean.dist, (aes(x=meanDistance, y=NumbSp, color=class, shape=class))) + theme_bw() +
  theme(axis.text=element_text(size=15), axis.title=element_text(size=15, face="bold"),
        legend.key.width = unit(0.8, units = "cm"), legend.text = element_text(size=15),
        legend.key=element_blank()) +
  ylab("number of species/record") +
  xlab("average distance (m)") +
  geom_point(size = 3, stroke = 0.6) +
  geom_smooth(method=lm, se=FALSE, fullrange=FALSE, size = 2) +
  scale_color_manual(name = "", labels = key.lab, values = plot.color) +
  scale_shape_manual(name = "", labels = key.lab, values = point.shape)

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```

```

#plot data with lumped species richness
p2 <- ggplot(prov.mean.dist, (aes(x=meanDistance, y=NumbSp, color=class, shape=class))) + theme_bw() +
  theme(axis.text=element_text(size=15), axis.title=element_text(size=15, face="bold"),
    legend.key.width = unit(0.8, units = "cm"), legend.text = element_text(size=15),
    legend.key=element_blank()) +
  ylab("number of species/record") +
  xlab("average distance (m)") +
  geom_point(size = 3, stroke = 0.6) +
  geom_smooth(method=lm, se=FALSE, fullrange=FALSE, size = 2) +
  scale_color_manual(name = "", labels = key.lab, values = plot.color) +
  scale_shape_manual(name = "", labels = key.lab, values = point.shape)

ggarrange(p1,p2, nrow = 1, ncol = 2, common.legend = TRUE, align = "hv", widths = 20, heights = 5)

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

  □ Amphibia ▲ Mammalia △ Actinopterygii ▽ Insecta
  ● Aves △ Reptilia △ Arachnida



```

```

tiff("sp_dist.tif", res=300, width = 7, height = 7, unit="in")
ggplot(mean.dist, (aes(x=meanDistance, y=NumbSp, color=class, shape=class))) + theme_bw() +
  theme(axis.text=element_text(size=15), axis.title=element_text(size=15, face="bold"),
    legend.key.width = unit(0.8, units = "cm"), legend.text = element_text(size=15),
    legend.key=element_blank(), legend.position = c(0.83, 0.85),
    legend.background=element_blank()) +
  ylab("Number of species record") +
  xlab("Average distance (m)") +
  geom_point(size = 3, stroke = 0.6) +
  geom_smooth(method=lm, se=FALSE, fullrange=FALSE, size = 2) +
  scale_color_manual(name = "", labels = key.lab, values = plot.color) +
  scale_shape_manual(name = "", labels = key.lab, values = point.shape)

## Warning: A numeric `legend.position` argument in `theme()` was deprecated in ggplot2
## 3.5.0.
## i Please use the `legend.position.inside` argument of `theme()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```

```

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

## pdf
## 2

Get number of conflict sites per province

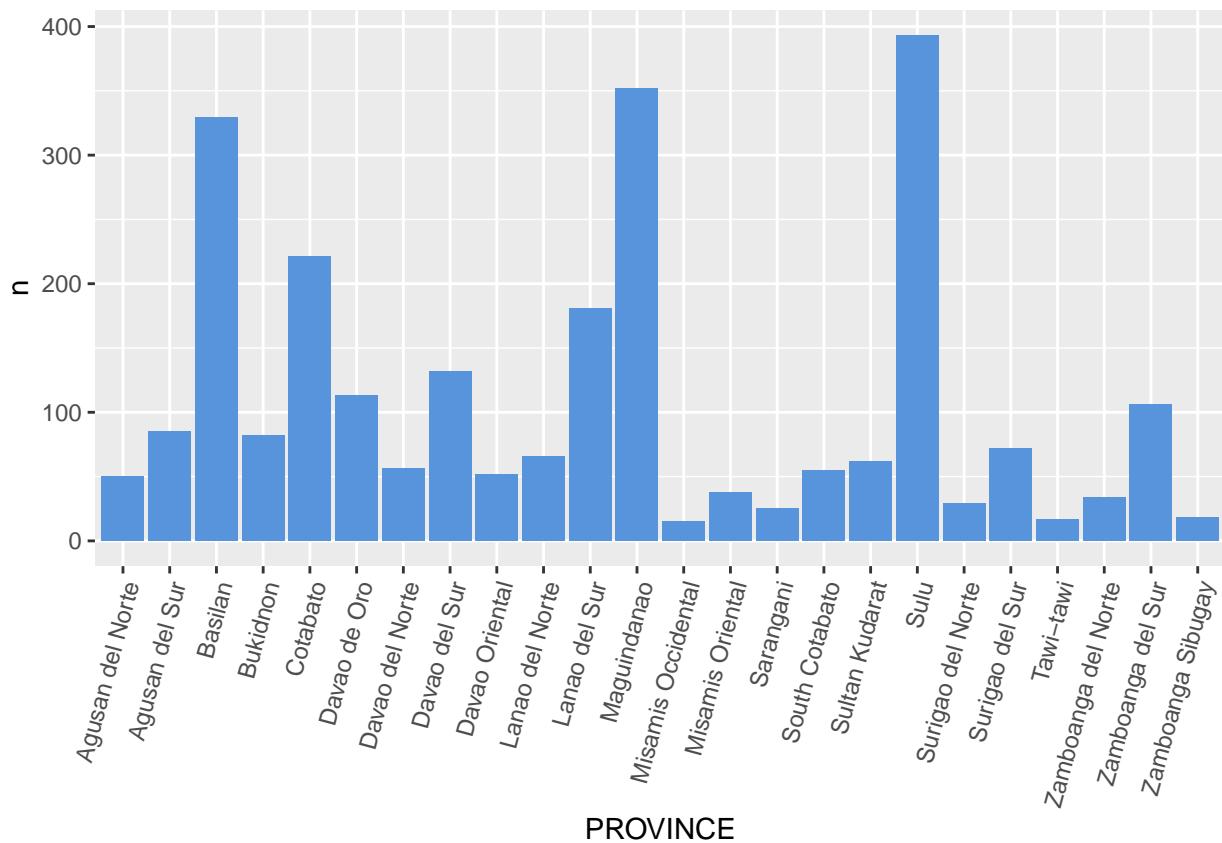
#we are not sure how the frequency of conflicts were scored in the paper as it was not mentioned so two ways can be considered here
#first, we can get frequency based on the number of unique conflict points within each province i.e., per point (unique latitude/longitude)
con.sub.2 <- con.sub
con.sub.2 <- con.sub.2 %>% filter(PROVINCE!="") #filter empty rows

con.freq <- con.sub.2 %>%
  group_by(PROVINCE) %>%
  mutate(ConflictFreq = n_distinct(latitude, longitude)) #use only the unique latitude
con.freq <- con.freq %>% distinct(PROVINCE, .keep_all = TRUE)
write.csv(con.freq, "conflict.frequency.unique.csv")

#second, count each row as a unique report of conflict regardless of coordinates
con.freq2 <- con.sub.2 %>%
  group_by(PROVINCE) %>%
  count()
write.csv(con.freq2, "conflict.frequency.rows.csv")

ggplot(data=con.freq2, aes(x=PROVINCE, y=n)) +
  geom_bar(stat="identity", fill="#5794db") +
  guides(x = guide_axis(angle = 75))

```



```
#combine biodiversity and conflict data (lumped)
prov.comb <- prov.lmp.dist %>% mutate(conflictFreq = if_else(county == "Agusan del Norte", 50,
  if_else(county == "Agusan del Sur", 85,
  if_else(county == "Basilan", 329,
  if_else(county == "Bukidnon", 82,
  if_else(county == "Cagayan de Oro", 0,
  if_else(county == "Camiguin Island", 0,
  if_else(county == "Cotabato", 221,
  if_else(county == "Davao", 0,
  if_else(county == "Davao Oriental", 52,
  if_else(county == "Davao de Oro", 113,
  if_else(county == "Davao del Norte", 56,
  if_else(county == "Davao del sur", 181,
  if_else(county == "Dinagat Island", 0,
  if_else(county == "General Santos", 0,
  if_else(county == "Lanao del Norte", 66,
  if_else(county == "Lanao del Sur", 181,
  if_else(county == "Maguindanao", 352,
  if_else(county == "Misamis Occidental", 15,
  if_else(county == "Misamis Oriental", 38,
  if_else(county == "North Cotabato", 0,
  if_else(county == "Sarangani", 25,
  if_else(county == "South Cotabato", 55,
  if_else(county == "Sultan Kudarat", 62,
  if_else(county == "Sulu", 393,
  if_else(county == "Surigao del Norte", 29,
  if_else(county == "Surigao del Sur", 72,
```

### Perform GLM

```

#glm using lumped species richness per province
#assign taxa as a factor
prov.comb$class <- factor(prov.comb$class)
#transform data
prov.comb$distlog <- log(prov.comb$lmpMeanDistance + 1)
prov.comb$freqlog <- log(prov.comb$conflictFreq + 1)

#plot data
h1 <- ggplot(prov.comb, aes(x = lmpMeanDistance, y = NumbSp)) +
    geom_point(pch = 1) + geom_smooth(method = lm)
h2 <- ggplot(prov.comb, aes(x = conflictFreq, y = NumbSp)) +

```

```

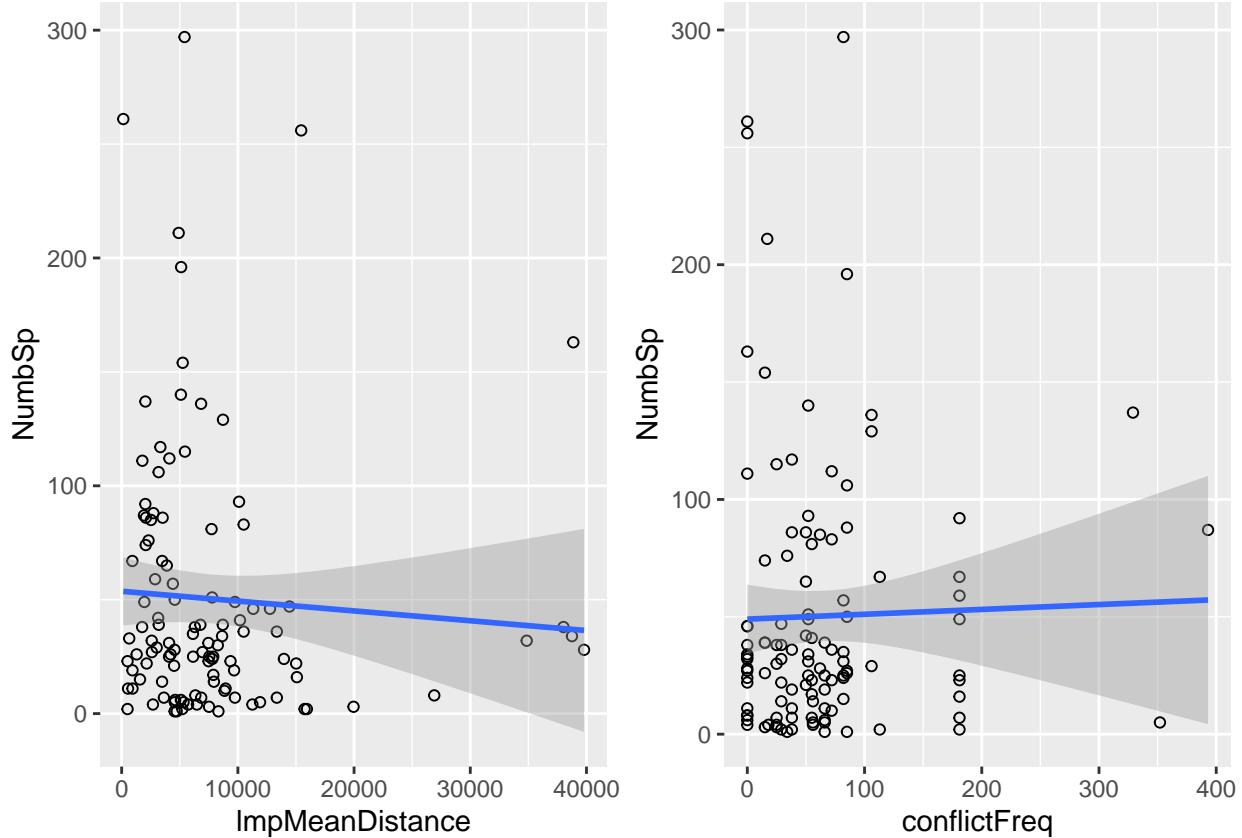
    geom_point(pch = 1) + geom_smooth(method = lm)
ggarrange(h1, h2, nrow = 1, ncol = 2, align = "hv")

```

```

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

```



```

#perform glm
glm.res.1 <- glm(NumbSp ~ freqlog + distlog + class,
                   family="poisson"(link="log"), data = prov.comb)
glm.res.2 <- glm(NumbSp ~ freqlog + class,
                   family="poisson"(link="log"), data = prov.comb)
glm.res.3 <- glm(NumbSp ~ distlog + class,
                   family="poisson"(link="log"), data = prov.comb)
glm.res.4 <- glm(NumbSp ~ freqlog + distlog,
                   family="poisson"(link="log"), data = prov.comb)

#compare glm results
AIC(glm.res.1, glm.res.2, glm.res.3, glm.res.4)

```

```

##          df      AIC
## glm.res.1 9 3608.833
## glm.res.2 8 3619.236
## glm.res.3 8 3651.504
## glm.res.4 3 6083.421
stepAIC(glm.res.1, direction='both') #use stepAIC function

```

```

## Start:  AIC=3608.83

```

```

## NumbSp ~ freqlog + distlog + class
##
##          Df Deviance    AIC
## <none>      3017.1 3608.8
## - distlog   1  3029.5 3619.2
## - freqlog   1  3061.7 3651.5
## - class     6  5503.7 6083.4
##
## Call: glm(formula = NumbSp ~ freqlog + distlog + class, family = poisson(link = "log"),
##           data = prov.comb)
##
## Coefficients:
## (Intercept)      freqlog      distlog  classAmphibia  classArachnida
## 3.55329        -0.05149       -0.04438      0.37876       0.46041
## classAves      classInsecta  classMammalia  classReptilia
## 1.32025         1.72873       -0.23824      0.78888
##
## Degrees of Freedom: 111 Total (i.e. Null); 103 Residual
## Null Deviance: 5683
## Residual Deviance: 3017 AIC: 3609
#get summary for best model
summary(glm.res.1)

##
## Call:
## glm(formula = NumbSp ~ freqlog + distlog + class, family = poisson(link = "log"),
##       data = prov.comb)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 3.553287  0.126761 28.031 < 2e-16 ***
## freqlog     -0.051487  0.007601 -6.773 1.26e-11 ***
## distlog     -0.044379  0.012537 -3.540  0.00040 ***
## classAmphibia 0.378760  0.078980  4.796 1.62e-06 ***
## classArachnida 0.460405  0.076193  6.043 1.52e-09 ***
## classAves    1.320248  0.066389 19.886 < 2e-16 ***
## classInsecta  1.728728  0.063862 27.070 < 2e-16 ***
## classMammalia -0.238237  0.083012 -2.870  0.00411 **
## classReptilia 0.788883  0.072272 10.915 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 5683.2 on 111 degrees of freedom
## Residual deviance: 3017.1 on 103 degrees of freedom
## AIC: 3608.8
##
## Number of Fisher Scoring iterations: 5
#glm with taxa as a random variable
#pred.scale <- scale(prov.comb[4:5]) #scale predictors
#prov.comb.s <- cbind(prov.comb[2:3], pred.scale)

```

```

glm.res.5 <- glmer(NumbSp ~ freqlog + distlog + (1|class),
                    family = "poisson"(link ="log"), data = prov.comb)
summary(glm.res.5)

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: NumbSp ~ freqlog + distlog + (1 | class)
## Data: prov.comb
##
##      AIC      BIC  logLik deviance df.resid
## 3644.6 3655.5 -1818.3   3636.6     108
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -9.9359 -3.1715 -0.6495  2.4200 19.6431
##
## Random effects:
## Groups Name        Variance Std.Dev.
## class  (Intercept) 0.4209   0.6487
## Number of obs: 112, groups: class, 7
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 4.190055  0.269402 15.553 < 2e-16 ***
## freqlog     -0.051447  0.007599 -6.770 1.28e-11 ***
## distlog     -0.044632  0.012535 -3.561  0.00037 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) freqlg
## freqlog -0.096
## distlog -0.399  0.010

#get R^2
rsquared(glm.res.5, method="trigamma")

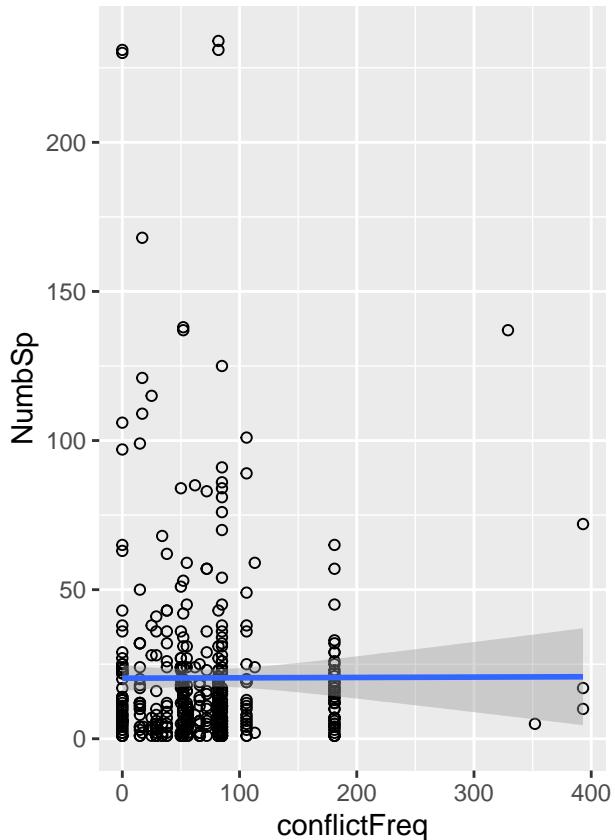
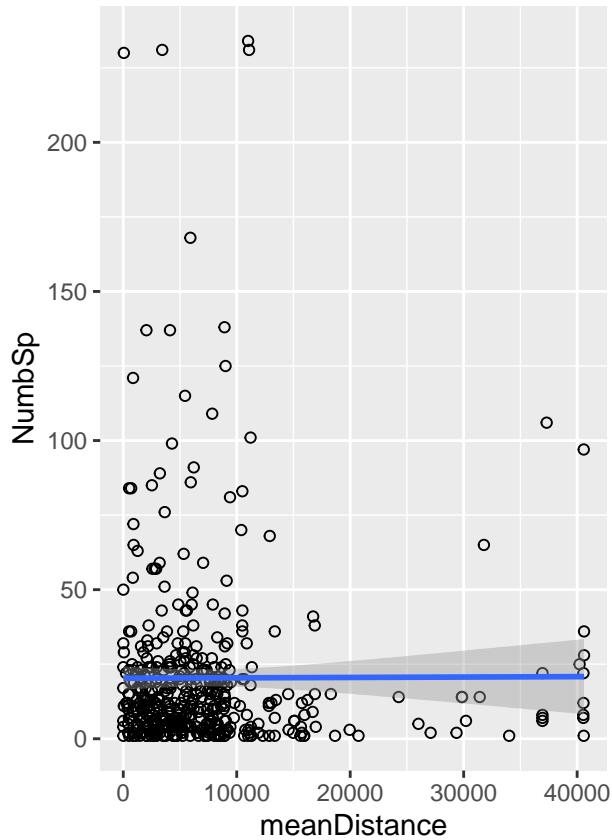
## Response family link method Marginal Conditional
## 1 NumbSp poisson log trigamma 0.01850072  0.9521265

#use data per point
#assign taxa as a factor
perpoint.comb$class <- factor(perpoint.comb$class)
#transform data
perpoint.comb$distlog <- log(perpoint.comb$meanDistance + 1)
perpoint.comb$freqlog <- log(perpoint.comb$conflictFreq + 1)

#plot data
h1 <- ggplot(perpoint.comb, aes(x = meanDistance, y = NumbSp)) +
  geom_point(pch = 1) + geom_smooth(method = lm)
h2 <- ggplot(perpoint.comb, aes(x = conflictFreq, y = NumbSp)) +
  geom_point(pch = 1) + geom_smooth(method = lm)
ggarrange(h1, h2, nrow = 1, ncol = 2, align = "hv")

```

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



```
#perform glm
glm.res.1 <- glm(NumbSp ~ freqlog + distlog + class,
                   family = "poisson"(link = "log"), data = perpoint.comb)
glm.res.2 <- glm(NumbSp ~ freqlog + class,
                   family = "poisson"(link = "log"), data = perpoint.comb)
glm.res.3 <- glm(NumbSp ~ distlog + class,
                   family = "poisson"(link = "log"), data = perpoint.comb)
glm.res.4 <- glm(NumbSp ~ freqlog + distlog,
                   family = "poisson"(link = "log"), data = perpoint.comb)

#compare glm results
AIC(glm.res.1, glm.res.2, glm.res.3, glm.res.4)
```

```
##             df      AIC
## glm.res.1  9 11538.32
## glm.res.2  8 11536.65
## glm.res.3  8 11656.29
## glm.res.4  3 14331.07
stepAIC(glm.res.1, direction = 'both') #use stepAIC function

## Start:  AIC=11538.32
## NumbSp ~ freqlog + distlog + class
##
##             Df Deviance   AIC
```

```

## - distlog 1 9615.0 11537
## <none> 9614.6 11538
## - freqlog 1 9734.6 11656
## - class 6 12419.4 14331
##
## Step: AIC=11536.65
## NumbSp ~ freqlog + class
##
##          Df Deviance AIC
## <none> 9615.0 11537
## + distlog 1 9614.6 11538
## - freqlog 1 9735.8 11656
## - class 6 12455.9 14366
##
## Call: glm(formula = NumbSp ~ freqlog + class, family = poisson(link = "log"),
##           data = perpoint.comb)
##
## Coefficients:
## (Intercept) freqlog classAmphibia classArachnida classAves
## 2.72365 -0.07196 0.32547 -0.05185 1.06010
## classInsecta classMammalia classReptilia
## 1.03821 -0.65465 0.41145
##
## Degrees of Freedom: 457 Total (i.e. Null); 450 Residual
## Null Deviance: 12510
## Residual Deviance: 9615 AIC: 11540
#get summary for best model
summary(glm.res.1)

##
## Call:
## glm(formula = NumbSp ~ freqlog + distlog + class, family = poisson(link = "log"),
##       data = perpoint.comb)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.759305  0.082424 33.477 < 2e-16 ***
## freqlog     -0.071763  0.006395 -11.222 < 2e-16 ***
## distlog     -0.004326  0.007521 -0.575   0.565
## classAmphibia 0.325910  0.059196  5.506 3.68e-08 ***
## classArachnida -0.051769  0.061466 -0.842   0.400
## classAves    1.059180  0.052478 20.183 < 2e-16 ***
## classInsecta  1.037147  0.051540 20.123 < 2e-16 ***
## classMammalia -0.654108  0.067277 -9.723 < 2e-16 ***
## classReptilia  0.412545  0.058763  7.021 2.21e-12 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 12508.2 on 457 degrees of freedom
## Residual deviance: 9614.6 on 449 degrees of freedom
## AIC: 11538

```

```

## Number of Fisher Scoring iterations: 6
#glm with taxa as a random variable
#pred.scale <- scale(perpoint.comb[6:7]) #scale predictors
#perpoint.comb.s <- cbind(perpoint.comb[2:3], pred.scale)

glm.res.5 <- glmer(NumbSp ~ freqlog + distlog + (1|class),
  family = "poisson"(link = "log"), data = perpoint.comb)
summary(glm.res.5)

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson  ( log )
## Formula: NumbSp ~ freqlog + distlog + (1 | class)
## Data: perpoint.comb
##
##      AIC      BIC  logLik deviance df.resid
##  11575.8  11592.3 -5783.9   11567.8      454
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -6.079 -2.965 -1.288  1.016 36.356
##
## Random effects:
## Groups Name        Variance Std.Dev.
## class  (Intercept) 0.3216   0.5671
## Number of obs: 458, groups: class, 7
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 3.064396  0.224596 13.644   <2e-16 ***
## freqlog     -0.071702  0.006393 -11.215   <2e-16 ***
## distlog     -0.004403  0.007520  -0.586    0.558
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) freqlg
## freqlog -0.082
## distlog -0.276 -0.056
#get R^2
rsquared(glm.res.5, method = "trigamma")

## Response family link  method Marginal Conditional
## 1  NumbSp poisson  log trigamma 0.03255409  0.8597707

```

Calculate the proportion of species occurrence records and conflict sites positioned in different land cover types

```

#here we calculate the number of sites found in different land cover types
#this demonstrates the artifact of sampling, which was not considered in the paper
#land cover data of 1988 was from Swedish Space Corporation (SSC 1988)
#retrieved from https://data.humdata.org/dataset/29a3760f-3170-4555-b5d7-1fdb6cfb5a69?force_layout=desk
#land cover of 2020 was from the Philippine National Mapping and Resource Information Authority
#data was retrieved from geoportal PH (https://www.geoportal.gov.ph/)

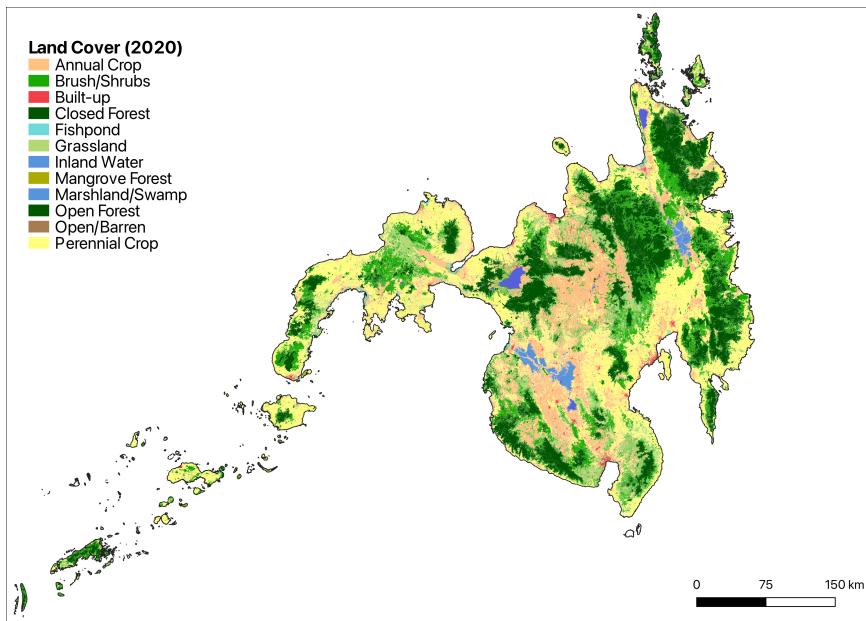
```

```

#we used land cover as a proxy for tree density and forest canopy data
#the national land cover data is ground-truthed
#this is only for simplistic illustration of sampling artifact
#the original land cover format is shapefile
#we converted it to raster file for fast data processing
#data conversion was done in QGIS using the plugin "rasterize"
#raster pixel size is 0.0008 or approximately 100 meters
#all raster files have the uniform coordinate reference system, resolution and extent

#here is the 2020 land cover of Mindanao
knitr:::include_graphics("lc_2020.png")

```



```

#read data
files <- list.files("land cover/", pattern = "tif", full.names=TRUE)
raster.files <- lapply(files, raster)

raster.files

## [[1]]
## class      : RasterLayer
## dimensions : 7356, 9030, 66424680  (nrow, ncol, ncell)
## resolution : 8e-04, 8e-04  (x, y)
## extent     : 119.381, 126.605, 4.586858, 10.47166  (xmin, xmax, ymin, ymax)
## crs        : +proj=longlat +datum=WGS84 +no_defs
## source     : 1988_lc_mindanao_rast100m.tif
## names      : X1988_lc_mindanao_rast100m
##
##
## [[2]]
## class      : RasterLayer
## dimensions : 7356, 9030, 66424680  (nrow, ncol, ncell)
## resolution : 8e-04, 8e-04  (x, y)
## extent     : 119.381, 126.605, 4.586858, 10.47166  (xmin, xmax, ymin, ymax)
## crs        : +proj=longlat +datum=WGS84 +no_defs

```

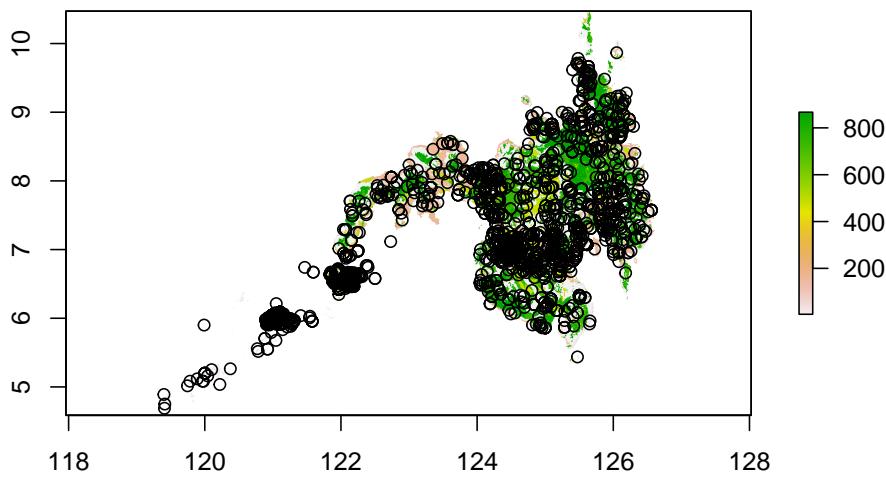
```

## source      : 2020_lc_mindanao_rast100m.tif
## names      : X2020_lc_mindanao_rast100m

#resample raster files to have a uniform dimension/resolution
standard <- raster.files[[1]]
raster.resamp <- list(standard)
for (i in 2:length(raster.files)) {
  raster.resamp[[i]] <- resample(raster.files[[i]], standard,
                                  method='bilinear')}
#stack raster files
lc <- stack(raster.files)

#plot data
plot(lc$X1988_lc_mindanao_rast100m)
points(con.sub$longitude, con.sub$latitude)

```



```

con.coor <- cbind(con.sub$longitude, con.sub$latitude)

#extract data from land cover using conflict data coordinates
con.lc.data <- extract(lc, con.coor)
colnames(con.lc.data)[1] <- "LC_1988"
colnames(con.lc.data)[2] <- "LC_2020"
con.lc.data <- as.data.frame(na.omit(con.lc.data))
head(con.lc.data)

```

```

##    LC_1988 LC_2020
## 1      756      5
## 2      154      4
## 3        5      4
## 4      154      9
## 5      154      2
## 6     867     11

```

```

#land cover types code
lc.88.d <- read.csv("1988_lc_mindanao_code.csv")
lc.88.d

```

code	category
4	Lake
5	Cultivated Area mixed with brushland/grassland
7	Built-up Area

```

## 4    13          Quarry
## 5    15          Unclassified
## 6    37          Coconut plantations
## 7    42          Open Canopy
## 8   154          Crop land mixed with coconut plantation
## 9   487          Arable land, crops mainly cereals and sugar
## 10  504          Coral Reef
## 11  557          Riverbeds
## 12  667          Siltation pattern in lake
## 13  683          Mangrove vegetation
## 14  706          Fishponds derived from mangrove
## 15  728          Marshy area and swamp
## 16  746          Closed Canopy
## 17  756          Crop land mixed with other plantation
## 18  776          Other plantations
## 19  840          Grassland, grass covering > 70 percent
## 20  867          Open Canopy

```

```

lc.20.d <- read.csv("2020_lc_mindanao_code.csv")
lc.20.d

```

```

##     code      category
## 1     1      Grassland
## 2     2      Annual Crop
## 3     3  Marshland/Swamp
## 4     4  Perennial Crop
## 5     5      Built-up
## 6     6      Fishpond
## 7     7      Inland Water
## 8     8  Mangrove Forest
## 9     9      Brush/Shrubs
## 10   10      Closed Forest
## 11   11      Open Forest
## 12   12      Open/Barren
## 13   13      No data

```

```

#add land cover descriptions
con.lc.data.m1 <- con.lc.data %>% mutate(class_1988 = if_else(LC_1988 == 4, "Lake",
  if_else(LC_1988 == 5, "Cultivated Area mixed with brushland/grassland",
  if_else(LC_1988 == 7, "Built-up Area",
  if_else(LC_1988 == 13, "Quarry",
  if_else(LC_1988 == 15, "Unclassified",
  if_else(LC_1988 == 37, "Coconut plantations",
  if_else(LC_1988 == 42, "Open Canopy",
  if_else(LC_1988 == 154, "Crop land mixed with coconut plantation",
  if_else(LC_1988 == 487, "Arable land, crops mainly cereals and sugar",
  if_else(LC_1988 == 504, "Coral Reef",
  if_else(LC_1988 == 557, "Riverbeds",
  if_else(LC_1988 == 667, "Open/Siltation pattern in lake",
  if_else(LC_1988 == 683, "Mangrove vegetation",
  if_else(LC_1988 == 706, "Fishponds derived from mangrove",
  if_else(LC_1988 == 728, "Marshy area and swamp",
  if_else(LC_1988 == 746, "Closed Canopy",
  if_else(LC_1988 == 756, "Crop land mixed with other plantation",
  if_else(LC_1988 == 776, "Other plantations",

```

```

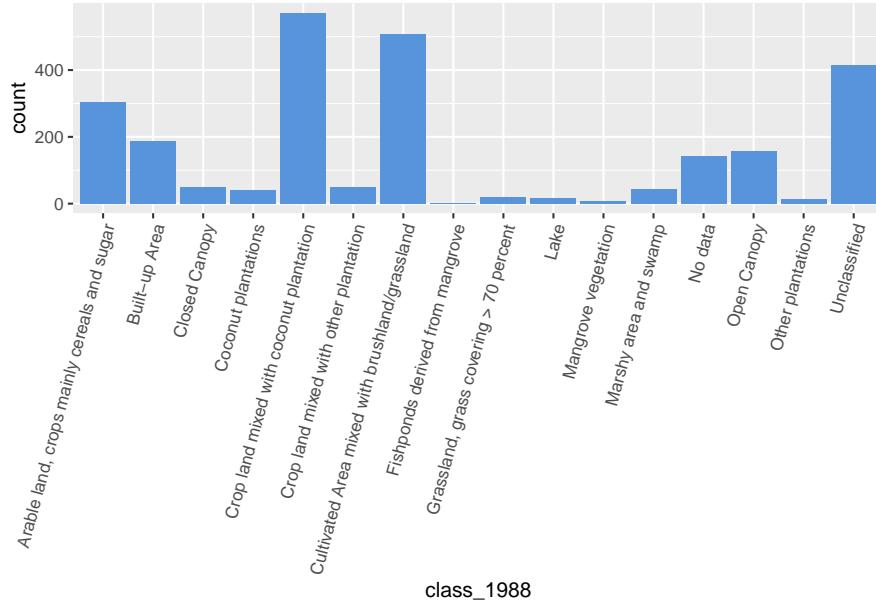
if_else(LC_1988 == 840, "Grassland, grass covering > 70 percent",
if_else(LC_1988 == 867, "Open Canopy", "No data"))))))))))))))))))))

con.lc.data.m2 <- con.lc.data %>% mutate(con.lc.data.m1, class_2020 = if_else(LC_2020 == 1, "Grassland"
if_else(LC_2020 == 2, "Annual Crop",
if_else(LC_2020 == 3, "Marshland/Swamp",
if_else(LC_2020 == 4, "Perennial Crop",
if_else(LC_2020 == 5, "Built-up",
if_else(LC_2020 == 6, "Fishpond",
if_else(LC_2020 == 7, "Inland Water",
if_else(LC_2020 == 8, "Mangrove Forest",
if_else(LC_2020 == 9, "Brush/Shrubs",
if_else(LC_2020 == 10, "Closed Forest",
if_else(LC_2020 == 11, "Open Forest",
if_else(LC_2020 == 12, "Open/Barren", "No data")))))))))))))

#count the number of conflict sites in each land cover type
#1988 data
con.lc.count.88 <- con.lc.data.m2 %>%
  group_by(class_1988) %>%
  summarise(n())
colnames(con.lc.count.88)[2] <- "count"

ggplot(data=con.lc.count.88, aes(x=class_1988, y=count)) +
  geom_bar(stat="identity", fill="#5794db") +
  guides(x = guide_axis(angle = 75))

```



```

con.lc.count.88 <- as.data.frame(con.lc.count.88)
write.csv(con.lc.count.88, "con.lc.count.88.csv")

#2020 data
con.lc.count.20 <- con.lc.data.m2 %>%
  group_by(class_2020) %>%
  summarise(n())

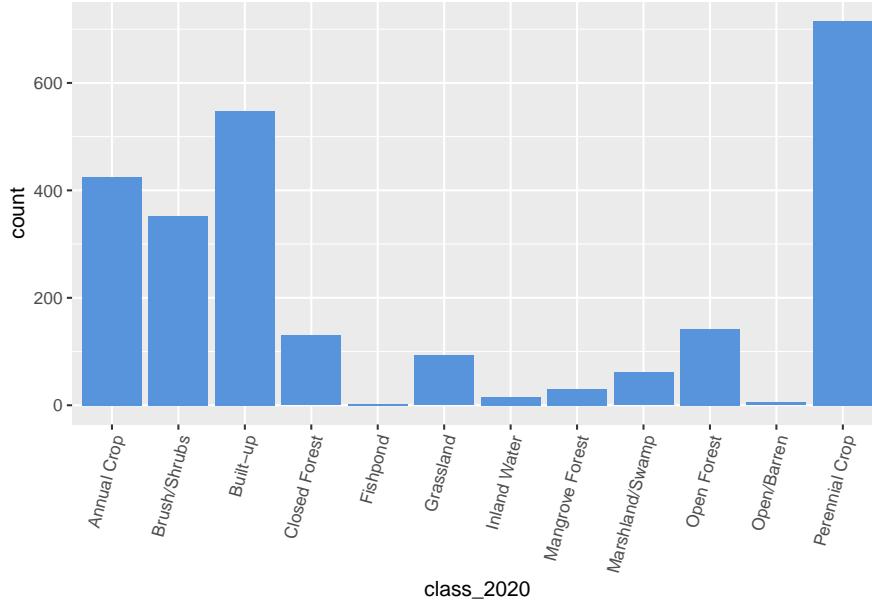
```

```

colnames(con.lc.count.20)[2] <- "count"

ggplot(data=con.lc.count.20, aes(x=class_2020, y=count)) +
  geom_bar(stat="identity", fill="#5794db") +
  guides(x = guide_axis(angle = 75))

```



```

con.lc.count.20 <- as.data.frame(con.lc.count.20)
write.csv(con.lc.count.20, "con.lc.count.20.csv")

#extract data from land cover using biodiversity data coordinates
bio.coor <- cbind(bio.sub$decimalLongitude, bio.sub$decimalLatitude)
bio.lc.data <- extract(lc, bio.coor)
colnames(bio.lc.data )[1] <- "LC_1988"
colnames(bio.lc.data )[2] <- "LC_2020"
bio.lc.data <- as.data.frame(na.omit(bio.lc.data ))
head(bio.lc.data)

##    LC_1988 LC_2020
## 1        5      1
## 2     154      4
## 3     154      4
## 4     154      4
## 5     154      4
## 6     154      4

#add land cover descriptions
bio.lc.data.m1 <- bio.lc.data %>% mutate(class_1988 = if_else(LC_1988 == 4, "Lake",
  if_else(LC_1988 == 5, "Cultivated Area mixed with brushland/grassland",
  if_else(LC_1988 == 7, "Built-up Area",
  if_else(LC_1988 == 13, "Quarry",
  if_else(LC_1988 == 15, "Unclassified",
  if_else(LC_1988 == 37, "Coconut plantations",
  if_else(LC_1988 == 42, "Open Canopy",
  if_else(LC_1988 == 154, "Crop land mixed with coconut plantation",
  if_else(LC_1988 == 487, "Arable land, crops mainly cereals and sugar",

```

```

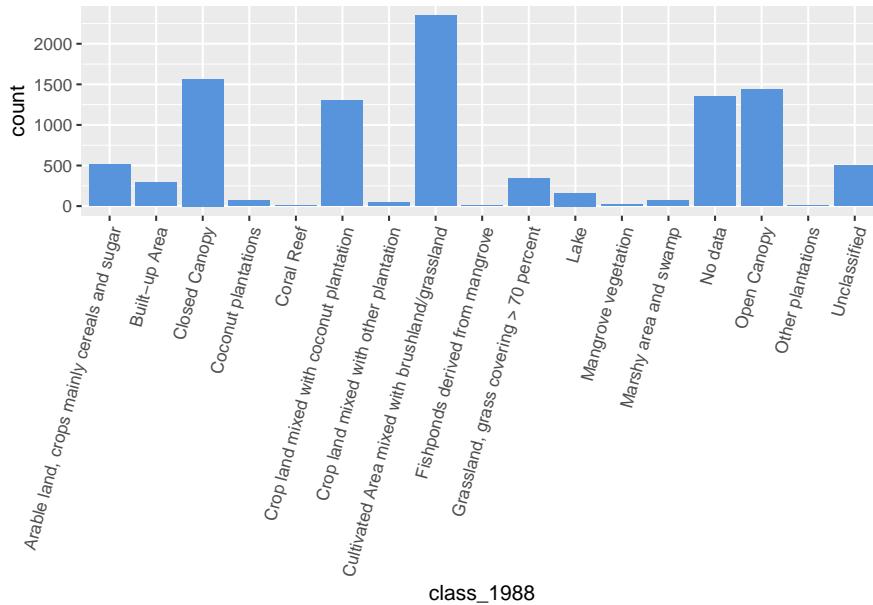
if_else(LC_1988 == 504, "Coral Reef",
if_else(LC_1988 == 557, "Riverbeds",
if_else(LC_1988 == 667, "Open/Siltation pattern in lake",
if_else(LC_1988 == 683, "Mangrove vegetation",
if_else(LC_1988 == 706, "Fishponds derived from mangrove",
if_else(LC_1988 == 728, "Marshy area and swamp",
if_else(LC_1988 == 746, "Closed Canopy",
if_else(LC_1988 == 756, "Crop land mixed with other plantation",
if_else(LC_1988 == 776, "Other plantations",
if_else(LC_1988 == 840, "Grassland, grass covering > 70 percent",
if_else(LC_1988 == 867, "Open Canopy", "No data"))))))))))))))))))))

bio.lc.data.m2<- bio.lc.data %>% mutate(bio.lc.data.m1, class_2020 = if_else(LC_2020 == 1, "Grassland",
if_else(LC_2020 == 2,"Annual Crop",
if_else(LC_2020 == 3, "Marshland/Swamp",
if_else(LC_2020 == 4, "Perennial Crop",
if_else(LC_2020 == 5, "Built-up",
if_else(LC_2020 == 6, "Fishpond",
if_else(LC_2020 == 7, "Inland Water",
if_else(LC_2020 == 8, "Mangrove Forest",
if_else(LC_2020 == 9, "Brush/Shrubs",
if_else(LC_2020 == 10, "Closed Forest",
if_else(LC_2020 == 11, "Open Forest",
if_else(LC_2020 == 12, "Open/Barren", "No data")))))))))))))

#count the number of conflict sites in each land cover type
#1988 data
bio.lc.count.88 <- bio.lc.data.m2 %>%
  group_by(class_1988) %>%
  summarise(n())
colnames(bio.lc.count.88)[2] <- "count"

ggplot(data=bio.lc.count.88, aes(x=class_1988, y=count)) +
  geom_bar(stat="identity", fill="#5794db") +
  guides(x = guide_axis(angle = 75))

```



```

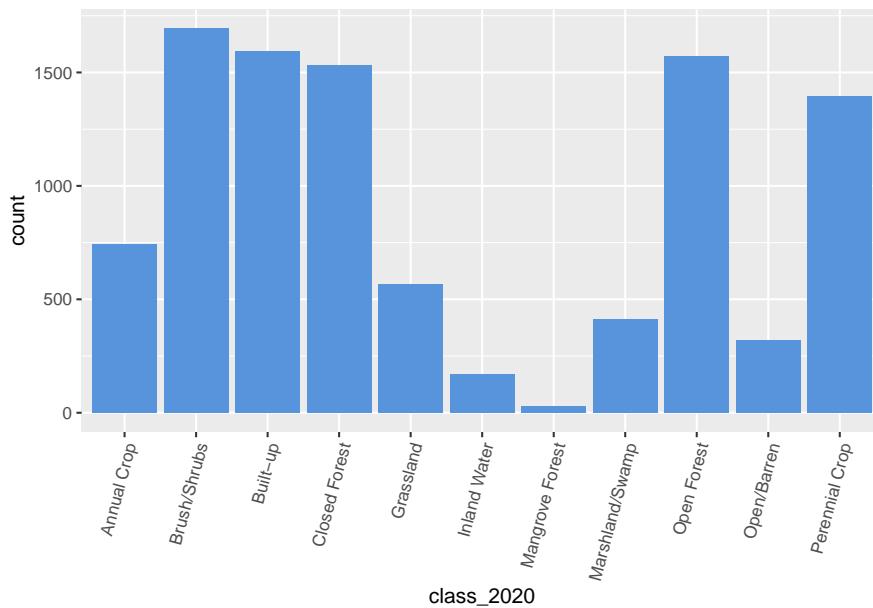
bio.lc.count.88 <- as.data.frame(bio.lc.count.88)
write.csv(bio.lc.count.88, "bio.lc.count.88.csv")

#2020 data
bio.lc.count.20 <- bio.lc.data.m2 %>%
  group_by(class_2020) %>%
  summarise(n())
colnames(bio.lc.count.20)[2] <- "count"

bio.lc.count.20 <- as.data.frame(bio.lc.count.20)

ggplot(data=bio.lc.count.20, aes(x=class_2020, y=count)) +
  geom_bar(stat="identity", fill="#5794db") +
  guides(x = guide_axis(angle = 75))

```



```
bio.lc.count.20 <- as.data.frame(bio.lc.count.20)
write.csv(bio.lc.count.20, "bio.lc.count.20.csv")
```