

```
#####
##### Selection of unions from zones for SE survey #####
##### NFI #####
##### Author: Rashed Jalal #####
##### contact: rashed.jalal@fao.org #####
#####

rm(list=ls())
# rootdir <- "Z:/4_GIS_Data/5_GIS_Data/All_GIS_DATA/296_SE_Survey_design/"
rootdir <- "C:/Users/Rashed/Desktop/296_SE_Survey_design_temp/"

setwd(rootdir)

library(xlsx)
library(foreign)
library(plyr)
library(sampling)

options(stringsAsFactors = FALSE)
options(scipen = 999)

#### RS and BNL ####
uni_lc <- read.dbf("GIS/Final_union_selection/GeoDB/SE_Unions_LC2015.dbf", as.is = T)
names(uni_lc)

uni_lc$Area_ha <- uni_lc$Shape_Area / 10000

length(unique(uni_lc$CODE_R))
length(unique(uni_lc$ID))

bnl_rs_area <- aggregate(uni_lc$Area_ha[uni_lc$CODE_R %in% c("BNL", "RS")] ~ uni_lc$ID[uni_lc$CODE_R %in% c("BNL", "RS")])
names(bnl_rs_area) <- c("ID", "bnl_rs_area_ha")
uni_area <- aggregate(uni_lc$Area_ha ~ uni_lc$ID, FUN = sum)
names(uni_area) <- c("ID", "Uni_area_ha")

uni_bnl_rs <- merge(uni_area, bnl_rs_area, by = "ID", all.x = T)
uni_bnl_rs <- arrange(uni_bnl_rs, ID)

nrow(uni_bnl_rs)

uni_bnl_rs$bnl_rs_prc <- 100 * uni_bnl_rs$bnl_rs_area_ha / uni_bnl_rs$Uni_area_ha
hist(uni_bnl_rs$bnl_rs_prc)

#### Read union dbf file and create unique ID ####
df1 <- read.dbf("GIS/All/Stratified/SE_Survey_variables_Union_stratified_proportional_BUTM2010.dbf")
names(df1)

## Data cleaning ##
# identical(df1$DISTNAME_x, df1$DISTNAME_y)
ncol(df1)
df1 <- df1[, !duplicated(t(df1))]

nrow(df1)
length(unique(df1$ID))
length(df1$uniq_code)

df1$Zone_x <- NULL
df1$Zone_y <- NULL
df1$THANAME_1 <- NULL
df1$UNINAME_1 <- NULL
df1$TC2000_y <- NULL
df1$TC2000_x <- NULL
df1$TC2014 <- NULL
df1$TC_change <- NULL
df1$TC_change_ <- NULL
df1$TC_HH_Area <- NULL
df1$strat_f <- NULL
df1$Selection <- NULL
```

```

names(df1) <- c("ID", "uniq_code", "DIST_NAME", "UPZ_NAME", "UNI_NAME", "DIV_NAME", "DIV_CODE", "DI
"UNI_CODE", "RL", "TC2000_prc", "TC2014_prc", "Fire_perc_", "Fire_percl", "Fire_per_l", "BU_prc
"Wt", "Zone", "HT", "HG", "HI", "HO", "PB", "PM", "PF", "SR", "Zone_SF_bf", "U_area_ha")

df1$uniq_code <- paste0(df1$DIV_NAME, df1$DIST_NAME, df1$UPZ_NAME, df1$UNI_NAME, df1$UNI_CODE)
# length(unique(df1$uniq_code))
# df1$uniq_code[duplicated(df1$uniq_code) == T]

#### Stratification ####
df1$TC2000_ha <- df1$TC2000_prc * df1$U_area_ha / 100
df1$TC2014_ha <- df1$TC2014_prc * df1$U_area_ha / 100
df1$TC_HH_Area <- df1$TC2014_ha/df1$U_area_ha/df1$HT # Stratification variable

df1$TC_HH_Area[df1$Zone == "Sundarbans"] <- NaN # has to be removed as there are no HH
df1$TC_HH_Area[is.infinite(df1$TC_HH_Area)] <- NaN

## Stratification
tempData <- lapply(split(df1, df1$Zone_SF_bf), function(x) {
  x$strata <- with(x, cut(TC_HH_Area, quantile(TC_HH_Area, na.rm = T),
    labels = 1:4, include.lowest = TRUE))
  x
})

dbf <- do.call(rbind, tempData)
dbf$strata <- paste0(dbf$Zone_SF_bf, "_", dbf$strata)
dbf <- arrange(dbf, strata)

dbf <- merge(dbf, uni_bnl_rs[,c("ID", "bnl_rs_area_ha", "bnl_rs_prc")], by = "ID", all.x = T)
dbf <- arrange(dbf, ID)
nrow(dbf)

##### Selection of union #####
## Sampling intensity per zone ##
# smpl_intsty <- aggregate(cbind(U_area_ha, HT) ~ Zone_SF_bf, data = dbf, FUN = sum)
# smpl_intsty$smpl_no <- ifelse(smpl_intsty$Zone_SF_bf == "Sundarbans", 1063, 1268)
# smpl_intsty$intensity <- 100 * smpl_intsty$smpl_no / smpl_intsty$HT

## Data cleaning
df2 <- dbf[!is.na(dbf$TC_HH_Area),] # Remove the NaN
nrow(df2)
df2 <- df2[!is.na(df2$bnl_rs_area_ha),] # Remove the NaN
nrow(df2)
df2$div_str <- paste0(df2$DIV_NAME, "_", df2$strata)

## Number of union per strata ##
sample_nos <- aggregate(TC_HH_Area ~ DIV_NAME + strata, data = df2, FUN = length)
names(sample_nos) <- c("DIV_NAME", "strata", "Uni_div_strata")

for (i in 1:nrow(sample_nos)) {
  sample_nos$Uni_strata[i] <- sum(sample_nos$Uni_div_strata[sample_nos$strata== sample_nos$strata[i]
}]

sample_nos$intensity <- 16 / sample_nos$Uni_strata
sample_nos$uni_sample <- ceiling(sample_nos$intensity * sample_nos$Uni_div_strata)
sample_nos$div_str <- paste0(sample_nos$DIV_NAME, "_", sample_nos$strata)

aggregate(sample_nos$uni_sample ~ sample_nos$strata, FUN = sum)

#####

# sample_nos <- aggregate(HT ~ strata, data = df2, FUN = sum)
# sample_nos$zone <- substr(sample_nos$strata, 1, nchar(sample_nos$strata)-2)
# # sample_nos <- merge(sample_nos, smpl_intsty[, c("Zone_SF_bf", "intensity")], by.x = "zone", by.y
# # sample_nos$hh_no_v1 <- ceiling(sample_nos$HT * sample_nos$intensity / 100)
# # sample_nos$uni_no <- ifelse(sample_nos$zone == "Coastal", 6, ifelse(sample_nos$zone == "Hill", 9

```

```

# #
# sample_nos$uni_no <- 16
# # sample_nos$hh_no_v2 <- sample_nos$uni_no * 20

## Data cleaning
# thrsld_area <- 30
# thrsld_area_uni <- 40
# temp <- df2[df2$bnl_rs_area_ha >= thrsld_area & df2$U_area_ha >=thrsld_area_uni,]
# nrow(temp)
# xxx <- aggregate(temp$bnl_rs_area_ha ~ temp$strata, FUN = length)
# xxx
# min(xxx$temp$bnl_rs_area_ha`)

prcntl <- 0.15
stlmnt_thrsld <- dply(df2, .(strata), function(x) quantile(x$bnl_rs_area_ha, c(prcntl)))
names(stlmnt_thrsld) <- c("strata", paste0("prcntl_",prcntl *100))

df2 <- merge(df2, stlmnt_thrsld, by = "strata", all.x = T)
df2 <- arrange(df2, strata)

# area_dist1 <- aggregate(df2$bnl_rs_area_ha ~ df2$strata, FUN = function(x) c(n = length(x), sm = s
#
#
#
# area_dist <- aggregate(df2$bnl_rs_area_ha ~ df2$strata, FUN = function(x) (avl = length(x) - length
#
# names(area_dist)
# area_dist
# sum(area_dist$`df2$bnl_rs_area_ha`)
# min(area_dist$`df2$bnl_rs_area_ha`)

df3 <- df2[df2$bnl_rs_area_ha > eval(parse(text = paste0("df2$prcntl_",prcntl *100))),]

# df2 <- df2[df2$U_area_ha > thrsld_area,]
nrow(df3)
aggregate(df3$bnl_rs_area_ha ~ df3$strata, FUN = length)

## Selection of union ##
sample_nos <- arrange(sample_nos, div_str)
df3 <- arrange(df3, div_str)
selection <- strata(df3,"div_str", sample_nos$uni_sample, method = "srswor")
selection$selection <- 1
nrow(selection)

df3$selection <- 0
df3[selection$ID_unit,"selection"] <- 1

df4 <- df3[df3$selection == 1,]
f_selection <- strata(df4,"strata", rep.int(16,20), method = "srswor")
nrow(f_selection)

df4$selection <- 0
df4[f_selection$ID_unit,"selection"] <- 1

# sample_nos <- arrange(sample_nos, strata)
# df3 <- arrange(df3, strata)
# selection <- strata(df3,"strata", sample_nos$uni_no, method = "srswor")
# selection$selection <- 1
# nrow(selection)
#
# df3[selection$ID_unit,"selection"] <- 1

## Incorporate in the dbf ##

```

```
dbf <- merge(x = dbf, y = df4[ , c("ID", "selection")], by = "ID", all.x=TRUE)
dbf <- arrange(dbf, ID)

dbf$selection[is.na(dbf$selection)] <- 0

aggregate(dbf$selection ~ dbf$strata, FUN = sum) # Check
aggregate(dbf$selection ~ dbf$Zone_SF_bf, FUN = sum) # Check
aggregate(dbf$selection ~ dbf$DIV_NAME, FUN = function(x) c(selected = sum(x), total = length(x), pr

#### Write the data ####

write.dbf(dbf, "GIS/Final_union_selection/SE_Unions_equal_BUTM2010_final.dbf")

write.xlsx(dbf, "GIS/Final_union_selection/Final_selected_equal_union.xlsx", row.names = F)

####
```