Data Cleaning logbook

Juan Carlos Villaseñor-Derbez

last compilation: 2019-02-20

Contents

The commit history for this file can be seen here

```r
suppressPackageStartupMessages({
  library(rfishbase)
  library(tidyverse)
})
```

Ecological data

Invertebrates

```r
invert <- read.csv(here::here("raw_data", "invert_data.csv"), strip.white = T, stringsAsFactors = F) %>%
  janitor::clean_names() %>%
  rename(generoespecie = genero_especie) %>%
  filter(comunidad %in% c("Isla Natividad", "Maria Elena", "Punta Herrero")) %>%
  filter(!is.na(rc)) %>%
  # filter(rc %in% c("La Plana / Las Cuevas - La Dulce / Babencho",
  #                  "Cabezo - Cabezo (Control)",
  #                  "Manchon - Manchon (Control)",
  #                  "El Faro - El Faro (Control)")) %>%
  # filter(!(rc %in% c("El Faro - El Faro (Control)") & ano == 2012)) %>%
  filter(between(abundancia, 0.1, 100)) %>%
  group_by(comunidad, ano, rc, sitio, zona, transecto, generoespecie) %>%
  summarize(abundancia = sum(abundancia)) %>%
  ungroup() %>%
  mutate(year = case_when(comunidad == "Isla Natividad" ~ ano - 2006,
                          comunidad == "Maria Elena" ~ ano - 2012,
                          TRUE ~ ano - 2013)) %>%
  select(comunidad, ano, year, rc, sitio, zona, transecto, generoespecie, abundancia)
```

Fish

Species biology

```r
species_bio <- MPAtools::species_bio %>%
  mutate(Genus = as.character(str_split(GeneroEspecie, pattern = " ", simplify = T)[,1]),
         Genus = str_replace(Genus, " ", "")) %>%
  drop_na(a, b)

needed_spp <- c("Embiotoca jacksoni",
                "Kyphosus incisor",
                "Kyphosus sectatrix",
                "Pterois volitans",
                "Scarus guacamaia",
                "Chaetodon capistratus",
                "Chaetodon capistratus",
                "Chaetodon striatus",
                "Calamus pennatula")

FB_update <- rfishbase::length_weight(needed_spp, fields = c("a", "b")) %>%
  select(GeneroEspecie = sciname, a, b) %>%
  mutate(GeneroEspecie = case_when(GeneroEspecie %in% c("Kyphosus sectatrix", "Kyphosus incisor") ~ "Ky
                                   GeneroEspecie %in% c("Chaetodon capistratus", "Chaetodon striatus") ~
                                   TRUE ~ GeneroEspecie)) %>%
  group_by(GeneroEspecie) %>%
  summarize_all(mean, na.rm = T) %>%
  ungroup()

manual_fishbase_bayesian <- tibble(
  GeneroEspecie = c("Chromis punctipinnis",
                    "Oxyjulis californica",
                    "Heterodontus francisci",
                    "Rhacochilus vacca",
                    "Squatina californica"),
  a = c(0.01514, 0.00447, 0.00355, 0.01950, 0.00741),
  b = c(2.99, 3.14,3.16, 3.02, 3.04)
)

needed_genus <- c("Mycteroperca",
                  "Chromis",
                  "Stegastes",
                  "Haemulon",
                  "Acanthurus",
                  "Sebastes",
                  "Lutjanus",
                  "Scarus")

GenusLevelBio <- species_bio %>%
  select(Genus, ag = a, bg = b) %>%
  group_by(Genus) %>%
  summarize_all(mean, na.rm = T) %>%
  ungroup()
```

```r
species_bio %<>%
  select(GeneroEspecie, a, b) %>%
  rbind(FB_update) %>%
  rbind(manual_fishbase_bayesian) %>%
  janitor::clean_names() %>%
    rename(generoespecie = genero_especie)
```

Clean data

```r
fish <- read.csv(here::here("raw_data", "fish_data.csv"), strip.white = T, stringsAsFactors = F) %>%
  janitor::clean_names() %>%
  rename(generoespecie = genero_especie) %>%
  filter(comunidad %in% c("Isla Natividad", "Maria Elena", "Punta Herrero")) %>%
  filter(!is.na(rc)) %>%
  # filter(rc %in% c("La Plana / Las Cuevas - La Dulce / Babencho",
  #                  "Cabezo - Cabezo (Control)",
  #                  "Manchon - Manchon (Control)",
  #                  "El Faro - El Faro (Control)")) %>%
  # filter(!(rc %in% c("El Faro - El Faro (Control)") & ano == 2012)) %>%
  filter(abundancia > 0) %>%
  mutate(generoespecie = case_when(generoespecie == "Cephalopholis cruentatus" ~ "Cephalopholis cruenta
                                   generoespecie == "Haemulon macrostomun" ~ "Haemulon macrostomum",
                                   generoespecie == "Xcochin spp" ~ "Balistes capriscus",
                                   generoespecie == "Pterois volitans/miles" ~ "Pterois volitans",
                                   generoespecie == "Mycteroperca spp" ~ "Mycteroperca",
                                   generoespecie == "Chromis spp" ~ "Chromis",
                                   generoespecie == "Damisela spp" ~ "Stegastes",
                                   generoespecie == "Haemulon spp" ~ "Haemulon",
                                   generoespecie == "Lancero spp" ~ "Acanthurus",
                                   generoespecie == "Kyphosus sectatrix/incisor" ~ "Kyphosus",
                                   generoespecie == "Sebastes spp" ~ "Sebastes",
                                   generoespecie == "Loro spp" ~ "Scarus",
                                   generoespecie == "Lutjanus spp" ~ "Lutjanus",
                                   generoespecie == "Mariposa spp" ~ "Chaetodon",
                                   generoespecie == "Pluma spp" ~ "Calamus pennatula",
                                   generoespecie == "Sparisoma aurolineatum" ~ "Sparisoma aurofrenatum"
                                   generoespecie == "Vieja spp" ~ "Bodianus rufus",
                                   TRUE ~ generoespecie)) %>%
  mutate(Genus = as.character(str_split(generoespecie, pattern = " ", simplify = T)[,1]),
         Genus = str_replace(Genus, " ", "")) %>%
  left_join(species_bio, by = "generoespecie") %>%
  left_join(GenusLevelBio, by = "Genus") %>%
  mutate(a = ifelse(is.na(a), ag, a),
         b = ifelse(is.na(b), bg, b)) %>%
  group_by(comunidad, ano, rc, sitio, zona, transecto, generoespecie, talla, a,b) %>%
  summarize(abundancia = sum(abundancia)) %>%
  ungroup() %>%
  mutate(year = case_when(comunidad == "Isla Natividad" ~ ano - 2006,
                          comunidad == "Maria Elena" ~ ano - 2012,
                          TRUE ~ ano - 2013))

fish <- fish %>%
```

```
    select(comunidad, ano, year, rc, sitio, zona, transecto, generoespecie, talla, abundancia, a, b)
```

Socioeconomic data

CPI

```
cpi <- read.csv(here::here("raw_data", "CPI.csv"), stringsAsFactors = F, strip.white = T)

cpi_2014 <- cpi$CPI[cpi$Ano == 2014]
```

Landings and revenue

```
cooperatives_t <- c("Scpp cozumel scl",
                    "Scpp jose maria azcorra scl",
                    "Scpp buzos y pescadores de la baja california scl")

cooperatives_c <- c("Scpp langosteros del caribe scl",
                    "Scpp vigia chico scl",
                    "Scpp pesc nacionales de abulon sc de rl",
                    "Scpp la purisima scl",
                    "Scpp bahia tortugas scl")

cooperatives <- c(cooperatives_t, cooperatives_c)
```

```
L_conapesca <- tibble(UnidadEconomica = c("buzos y pescadores de la baja california",
                                          "bahia tortugas",
                                          "pesc nacionales de abulon",
                                          "la purisima",
                                          "cozumel",
                                          "jose maria azcorra",
                                          "vigia chico",
                                          "langosteros del caribe"),
                      Comunidad = c("Isla Natividad",
                                    NA, NA, NA,
                                    "Maria Elena",
                                    "Punta Herrero",
                                    NA, NA),
                      Zona = c("PN", "PN", "PN", "PN", "MC", "MC", "MC", "MC"),
                      Start = c(2006, 2006, 2006, 2006, 2012, 2013, 2012, 2012))
```

```
conapesca <- readRDS(here::here("raw_data","conapesca.rds")) %>%
  mutate(UnidadEconomica = as.character(UnidadEconomica),
         LugarDeCaptura = as.character(LugarDeCaptura)) %>%
  mutate(UnidadEconomica = ifelse(UnidadEconomica == "Scpp bahia tortugas s.c de r.l",
                                  "Scpp bahia tortugas scl",
                                  UnidadEconomica)) %>%
  filter(UnidadEconomica %in% cooperatives) %>%
  filter(NombreComun %in% c("Langosta roja ent. fca.", "Langosta ent. fca.")) %>%
  filter(!(UnidadEconomica == "Scpp cozumel scl" & !LugarDeCaptura %in% c("Bahia espiritu santo",
                                                                          "Bahia del del espiritu santo
```

```r
                                                      "Bahia espiritu santo"))) %>%
    group_by(UnidadEconomica, Ano, NombreComun) %>%
    summarize(PesoDesembarcado = sum(PesoDesembarcado),
              PesoVivo = sum(PesoVivo),
              Valor = sum(Valor)) %>%
    ungroup() %>%
    filter(!(UnidadEconomica == "Scpp buzos y pescadores de la baja california scl" & NombreComun == "Lang
    left_join(cpi, by = "Ano") %>%
    mutate(Valor = Valor * (CPI / cpi_2014),
           Group = case_when(UnidadEconomica %in% cooperatives_t ~ "Treated",
                             TRUE ~ "Control"),
           UnidadEconomica = gsub(pattern = "Scpp ", replacement = "", x = UnidadEconomica),
           UnidadEconomica = gsub(pattern = " scl", replacement = "", x = UnidadEconomica),
           UnidadEconomica = gsub(pattern = " s.c de r.l", replacement = "", x = UnidadEconomica),
           UnidadEconomica = gsub(pattern = " sc de rl", replacement = "", x = UnidadEconomica)) %>%
    left_join(L_conapesca, by = "UnidadEconomica") %>%
    mutate(Post = ifelse(Ano <= Start, 0, 1),
           years_centered = Ano - Start) %>%
    select(Zona, Grupo = Group, Comunidad, UnidadEconomica, Ano, Post, years_centered, NombreComun, PesoVi
    janitor::clean_names() %>%
    filter(!(zona == "MC" & years_centered %in% c(-3, 2))) %>%
    filter(years_centered > -7)

write.csv(x = invert, file = here::here("data", "invertebrates.csv"), row.names = F)

write.csv(x = fish, file = here::here("data", "fish.csv"), row.names = F)

write.csv(x = conapesca, file = here::here("data", "conapesca.csv"), row.names = F)
```