

Annual sampling density Figure 4

Diego Ellis Soto

2026-01-14

Creating annual sampling density plot

```
require(tidyverse)

## Loading required package: tidyverse

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

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.1     v stringr   1.6.0
## v ggplot2   4.0.1     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr    1.3.1
## v purrr    1.2.0

## -- 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

# require(plyr)
require(dplyr)
require(sf)

## Loading required package: sf
## Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE

require(ggplot2)

# redlining colors
holc_pal <- c('#92BC6B' # green
              , '#92C7C9' # blue
              , '#E7DC6B' # yellow
              , '#E47D67' # red
              , '#A9A9A9' #,
) # dark gray
#
```

```

# holc_area <- suppressWarnings(sf::st_read('../indir/holc_shp', quiet = TRUE) %>%
#     sf::st_cast('POLYGON') %>% # IMPORTANT
#     dplyr::filter(!sf::st_is_empty(.)) %>%
#     sf::st_make_valid() %>%
#     tibble::rowid_to_column() %>%
#     dplyr::mutate( id = paste(state, city, holc_id, holc_grade, rowid, sep = '_')
#                 , city_state = paste0(city, ' ', state)
#                 , area_holc_km2 = as.double(sf::st_area(.) / 1e+6)) %>%
#     dplyr::select(id, state, city, holc_id, holc_grade, city_state, area_holc_km2) )
#
# holc_area <- holc_area %>% dplyr::select(city, holc_grade, area_holc_km2) %>% dplyr::group_by(holc_grade)
#
holc_area = read.csv(
    '../indir/Biodiv_Greeness_Social/main_combined_2022-05-27.csv') |>
group_by(holc_grade) |>
dplyr::summarise(area_sum = sum(area_holc_km2)) %>%
dplyr::filter(holc_grade != 'E') %>%
as_tibble()
#
# setnames(holc_area, 'sum_area_holc_km2', 'area_sum')
names(holc_area) <- c('holc_grade', 'area_sum')

temporal_2000_2020 = read.table('../indir/Biodiv_Greeness_Social/R1_biodiv_col_code_by_holc_id_2000_2020.txt')
names(temporal_2000_2020) <- c('Type', 'Sum', 'holc_polygon_id')
temporal_2000_2020$holc_grade = substr(sub('.*?_', "", (sub(".*?", "", sub(".*?", "", temporal_2000_2020$Type))), 1), 1, 1)
temporal_2000_2020 = temporal_2000_2020 %>% filter(holc_grade %in% c('A', 'B', 'C', 'D'))
# A few HOLC polygons do not contain any bird observations from 2000-2020 which makes total sense
temporal_2000_2020 %>% dplyr::filter(Sum > 0) %>% summarise(length(unique(holc_polygon_id)))

## length(unique(holc_polygon_id))
## 1 8434

sum(temporal_2000_2020$Sum) # Most of bird biodiversity data in these cities was collected from 2000-2020
## [1] 9563644

# Load 1933-2022 data
temporal_trend = read.table('../indir/Biodiv_Greeness_Social/R1_biodiv_trend_by_time_holc_id_1933_2022.txt')
sum(temporal_trend$N_samples)

## [1] 12297506

names(temporal_trend) <- c('Year', 'holc_grade', 'Sum')
temporal_trend = temporal_trend %>% filter(holc_grade != 'E')

sum(temporal_2000_2020$Sum,na.rm=T) / sum(temporal_trend$Sum,na.rm=T) # 77.8 % of biodiversity data collected in HOLC areas

## [1] 0.7781461

```

```

temporal_all_data <- temporal_trend %>%
  group_by(holc_grade, Year) %>%
  dplyr::summarise(
    n_obs = sum(Sum, na.rm = TRUE),
    .groups = "drop"
  )

tmpppp = temporal_all_data %>% group_by(holc_grade, Year) # %>% mutate(cumsum = cumsum(n_obs))

trend_A = tmpppp %>% filter(holc_grade == 'A') %>%
  dplyr::mutate(cumsum_n_obs = cumsum(n_obs)) %>%
  left_join(holc_area) %>%
  dplyr::mutate(sampling_density = cumsum_n_obs /area_sum )

## Joining with 'by = join_by(holc_grade)'

trend_B = tmpppp %>% filter(holc_grade == 'B') %>% dplyr::mutate(cumsum_n_obs = cumsum(n_obs)) %>% left

## Joining with 'by = join_by(holc_grade)'

trend_C = tmpppp %>% filter(holc_grade == 'C') %>% dplyr::mutate(cumsum_n_obs = cumsum(n_obs)) %>% left

## Joining with 'by = join_by(holc_grade)'

trend_D = tmpppp %>% filter(holc_grade == 'D') %>% dplyr::mutate(cumsum_n_obs = cumsum(n_obs)) %>% left

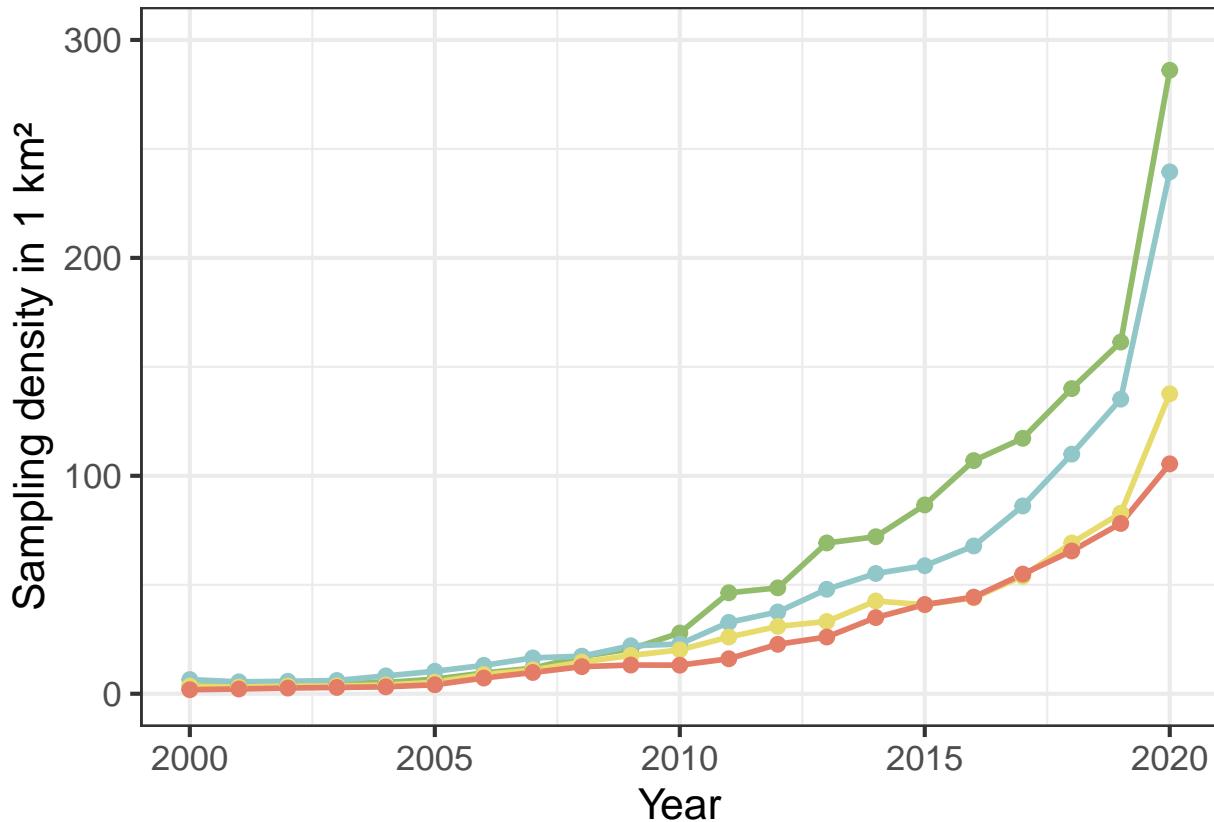
## Joining with 'by = join_by(holc_grade)'

temporal_all_data = rbind(trend_A,trend_B,trend_C,trend_D)

updated_figure_4 <- temporal_all_data %>%
  filter(Year >= 2000 & Year <= 2020) %>%
  ggplot(
    aes(
      x = Year,
      y = sampling_density,
      color = holc_grade,
      group = holc_grade
    )
  ) +
  geom_line(linewidth = 1) +
  geom_point(size = 2) +
  scale_color_manual(values = holc_pal) +
  theme_bw(16) +
  theme(legend.position = "none") +
  ylab("Sampling density in 1 km2") +
  coord_cartesian(ylim = c(0, 300)) + scale_y_continuous(breaks = seq(0, 300, by = 100))

updated_figure_4

```



```

ggsave('.../outdir/updated_figure_4.png'
      , width = 4.42
      , height = 5
      , dpi = 600
)

```

Recreating Figure 4 with HOLC area from originally deposited data (we note the replicator also uses both area estimates)

```

## Loading required package: mgcv

## Loading required package: nlme

## 
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':
## 
##     collapse

## This is mgcv 1.9-1. For overview type 'help("mgcv-package")'.

## Loading required package: sjPlot

```

```
##  
## Attaching package: 'sjPlot'  
  
## The following object is masked from 'package:ggplot2':  
##  
##     set_theme
```

Annual log sampling density

Annual sampling density

Predictors

Estimates

CI

p

Estimates

CI

p

(Intercept)

-451.53

-475.70 – -427.37

<0.001

-19867.51

-23799.40 – -15935.62

<0.001

Year

0.23

0.21 – 0.24

<0.001

9.91

7.96 – 11.87

<0.001

holc grade [B]

90.89

56.72 – 125.07

<0.001

4760.64

-799.89 – 10321.17

0.092

holc grade [C]

74.51

40.34 – 108.68

<0.001

10504.71

4944.18 – 16065.24

<0.001

holc grade [D]

44.77

10.60 – 78.94

0.011

11537.94

5977.41 – 17098.47

<0.001

Year × holc grade [B]

-0.05

-0.06 – -0.03

<0.001

-2.37

-5.14 – 0.39

0.091

Year × holc grade [C]

-0.04

-0.05 – -0.02

<0.001

-5.24

-8.01 – -2.47

<0.001

Year × holc grade [D]

-0.02

-0.04 – -0.01

0.010

-5.76

-8.52 – -2.99

<0.001

Observations

84

84

R2

0.982

0.720

AIC

454.261

803.248

```
## R version 4.4.1 (2024-06-14)
## Platform: aarch64-apple-darwin20
## Running under: macOS Sonoma 14.6
##
## Matrix products: default
## BLAS:    /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
## LAPACK:  /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib;  LAPACK v
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/Los_Angeles
## tzcode source: internal
##
## attached base packages:
## [1] stats      graphics   grDevices utils      datasets   methods    base
##
## other attached packages:
## [1] sjPlot_2.8.17 mgcv_1.9-1     nlme_3.1-167    sf_1.0-21
## [5] lubridate_1.9.4 forcats_1.0.1  stringr_1.6.0   dplyr_1.1.4
## [9] purrrr_1.2.0  readr_2.1.5    tidyrr_1.3.1   tibble_3.3.0
## [13] ggplot2_4.0.1 tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] gtable_0.3.6      bayestestR_0.15.2   xfun_0.52       insight_1.0.2
## [5] lattice_0.22-6    tzdb_0.5.0        vctrs_0.6.5     tools_4.4.1
## [9] sjstats_0.19.0    generics_0.1.4    datawizard_1.0.0 proxy_0.4-27
## [13] pkgconfig_2.0.3   Matrix_1.7-2      KernSmooth_2.23-26 RColorBrewer_1.1-3
## [17] S7_0.2.1         ggeffects_2.2.0   lifecycle_1.0.4  compiler_4.4.1
## [21] farver_2.1.2     sjmisc_2.8.10    textshaping_1.0.1 htmltools_0.5.8.1
## [25] class_7.3-23    yaml_2.3.10     pillar_1.11.1   classInt_0.4-11
## [29] tidyselect_1.2.1  sjlabelled_1.2.0  digest_0.6.37   performance_0.13.0
## [33] stringi_1.8.7    labeling_0.4.3    splines_4.4.1   fastmap_1.2.0
## [37] grid_4.4.1        cli_3.6.5        magrittr_2.0.4   dichromat_2.0-0.1
## [41] e1071_1.7-16    withr_3.0.2      scales_1.4.0    timechange_0.3.0
## [45] rmarkdown_2.29    ragg_1.4.0       hms_1.1.3       evaluate_1.0.3
## [49] knitr_1.50       parameters_0.24.1 rlang_1.1.6     Rcpp_1.1.0
## [53] glue_1.8.0        DBI_1.2.3        effectsize_1.0.0 rstudioapi_0.17.1
## [57] R6_2.6.1         systemfonts_1.2.3 units_0.8-7
```