

```
library(here)
```

```
## here() starts at /Users/ruanvanmazijk/Desktop/Cape-vs-SWA
```

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse
```

```
## v ggplot2 3.2.1    v purrr  0.3.3
## v tibble  2.1.3    v dplyr  0.8.3
## v tidyr   1.0.0    v stringr 1.4.0
## v readr   1.3.1    v forcats 0.4.0
```

```
## -- Conflicts ----- tidyverse_conflict
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(magrittr)
```

```
##
```

```
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
```

```
##
```

```
##      set_names
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
##      extract
```

```
library(raster)
```

```
## Loading required package: sp
```

```
##
```

```
## Attaching package: 'raster'
```

```
## The following object is masked from 'package:magrittr':
```

```
##
```

```
##      extract
```

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

```
##
```

```
##      select
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
##      extract
```

```
library(rgdal)
```

```
## rgdal: version: 1.4-7, (SVN revision 845)
```

```
## Geospatial Data Abstraction Library extensions to R successfully loaded
```

```
## Loaded GDAL runtime: GDAL 2.4.2, released 2019/06/28
```

```
## Path to GDAL shared files: /Library/Frameworks/R.framework/Versions/3.6/Resources/library/rgdal/gdal
```

```
## GDAL binary built with GEOS: FALSE
```

```
## Loaded PROJ.4 runtime: Rel. 5.2.0, September 15th, 2018, [PJ_VERSION: 520]
```

```
## Path to PROJ.4 shared files: /Library/Frameworks/R.framework/Versions/3.6/Resources/library/rgdal/proj
```

```
## Linking to sp version: 1.3-1
```

```

library(vegan)

## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.5-6

data <- read_csv(here(
  "data/derived-data/May-2019",
  "data-QDS.csv"
))

## Parsed with column specification:
## cols(
##   Elevation = col_double(),
##   MAP = col_double(),
##   PDQ = col_double(),
##   Surface_T = col_double(),
##   NDVI = col_double(),
##   CEC = col_double(),
##   Clay = col_double(),
##   Soil_C = col_double(),
##   pH = col_double(),
##   region = col_character(),
##   PC1 = col_double(),
##   lon = col_double(),
##   lat = col_double(),
##   QDS = col_character(),
##   n_EDS_in_region = col_double(),
##   n_collections = col_double(),
##   QDS_richness = col_double()
## )

GCFR_species <- read_csv(here(
  "data/derived-data/May-2019",
  "GCFR-species.csv"
))

## Parsed with column specification:
## cols(
##   species = col_character(),
##   n_collections = col_double()
## )

SWAFR_species <- read_csv(here(
  "data/derived-data/May-2019",
  "SWAFR-species.csv"
))

## Parsed with column specification:
## cols(
##   species = col_character(),
##   n_collections = col_double()
## )

GCFR_box <- readOGR(here("data/derived-data/borders/GCFR_box"))

```

```

## OGR data source with driver: ESRI Shapefile
## Source: "/Users/ruanvanmazijk/Desktop/Cape-vs-SWA/data/derived-data/borders/GCFR_box", layer: "value"
## with 1 features
## It has 1 fields

SWAFR_box <- readOGR(here("data/derived-data/borders/SWAFR_box"))

## OGR data source with driver: ESRI Shapefile
## Source: "/Users/ruanvanmazijk/Desktop/Cape-vs-SWA/data/derived-data/borders/SWAFR_box", layer: "value"
## with 1 features
## It has 1 fields

ZA_EDS <- readOGR(here("data/raw-data/QDGC/qdgc_zaf"), layer = "qdgc_03_zaf")

## OGR data source with driver: ESRI Shapefile
## Source: "/Users/ruanvanmazijk/Desktop/Cape-vs-SWA/data/raw-data/QDGC/qdgc_zaf", layer: "qdgc_03_zaf"
## with 14144 features
## It has 4 fields

AU_EDS <- readOGR(here("data/raw-data/QDGC/qdgc_au"), layer = "qdgc_03_au")

## OGR data source with driver: ESRI Shapefile
## Source: "/Users/ruanvanmazijk/Desktop/Cape-vs-SWA/data/raw-data/QDGC/qdgc_au", layer: "qdgc_03_au"
## with 89216 features
## It has 4 fields

GCFR_EDS <- crop(ZA_EDS, GCFR_box)
SWAFR_EDS <- crop(AU_EDS, SWAFR_box)

Larsen_grid <- rbind(GCFR_EDS, SWAFR_EDS)

Larsen_grid$edgc <- Larsen_grid$qdgc
Larsen_grid$qdgc <- str_remove(Larsen_grid$edgc, ".$")
Larsen_grid$hdgc <- str_remove(Larsen_grid$qdgc, ".$")
Larsen_grid$dgc <- str_remove(Larsen_grid$hdgc, ".$")

make_SpatialPointsDataFrame <- function(df) {
  SpatialPointsDataFrame(
    coords = df[, c("decimallongitude", "decimallatitude")],
    data = df[, "species"],
    proj4string = crs(Larsen_grid)
  )
}

GCFR_species_occ <- make_SpatialPointsDataFrame(read_csv(here(
  "data/derived-data/flora",
  "GCFR_clean_flora_2017-09-14.csv"
)))

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   family = col_character(),
##   genus = col_character(),
##   species = col_character(),
##   infraspecific epithet = col_character(),
##   scientificname = col_character(),

```

```

##   taxonrank = col_character(),
##   decimallatitude = col_double(),
##   decimallongitude = col_double(),
##   coordinateuncertaintyinmeters = col_double(),
##   coordinateprecision = col_logical()
## )

## Warning: 1 parsing failure.
##   row          col          expected actual
## 392500 coordinateprecision 1/0/T/F/TRUE/FALSE    0.0 '/Users/ruanvanmazijk/Desktop/Cape-vs-SWA/data/
SWAFR_species_occ <- make_SpatialPointsDataFrame(read_csv(here(
  "data/derived-data/flora",
  "SWAFR_clean_flora_2017-09-14.csv"
)))

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   family = col_character(),
##   genus = col_character(),
##   species = col_character(),
##   infraspecific epithet = col_character(),
##   scientificname = col_character(),
##   taxonrank = col_character(),
##   decimallatitude = col_double(),
##   decimallongitude = col_double(),
##   coordinateuncertaintyinmeters = col_double(),
##   coordinateprecision = col_logical()
## )

## Warning: 4 parsing failures.
##   row          col          expected actual
## 72178 coordinateprecision 1/0/T/F/TRUE/FALSE    20.0 '/Users/ruanvanmazijk/Desktop/Cape-vs-SWA/data/
## 126045 coordinateprecision 1/0/T/F/TRUE/FALSE    20.0 '/Users/ruanvanmazijk/Desktop/Cape-vs-SWA/data/
## 126046 coordinateprecision 1/0/T/F/TRUE/FALSE    10.0 '/Users/ruanvanmazijk/Desktop/Cape-vs-SWA/data/
## 225575 coordinateprecision 1/0/T/F/TRUE/FALSE    15.0 '/Users/ruanvanmazijk/Desktop/Cape-vs-SWA/data/
species_occ <- rbind(GCFR_species_occ, SWAFR_species_occ)

species_occ$EDS <- species_occ %>%
  Larsen_grid %>%
  pull(edgc)
species_occ@data$EDS %<>% as.character()
species_occ$QDS <- str_remove(species_occ$EDS, ".$")
species_occ$HDS <- str_remove(species_occ$QDS, ".$")
species_occ$DS <- str_remove(species_occ$HDS, ".$")

species_occ2 <- species_occ@data %>%
  mutate(
    lon = DS %>%
      str_extract("E\\d\\d\\d\\d") %>%
      str_remove("E") %>%
      as.numeric(),
    region = ifelse(lon >= 112, "SWAFR", "GCFR")
  )

```

```

) %>%
  split(.$region) %>%
  map(dplyr::select, -EDS, -lon, -region) %>%
  map(distinct)

GCFR_species %<>% filter(n_collections >= 5)
SWAFR_species %<>% filter(n_collections >= 5)

cells <- sort(unique(species_occ2$GCFR$QDS))
cells <- cells[cells %in% data$QDS]
species <- sort(unique(species_occ2$GCFR$species))
species <- species[species %in% GCFR_species$species]
n_cells <- length(cells)
n_species <- length(species)
GCFR_matrix <- matrix(nrow = n_cells, ncol = n_species)
rownames(GCFR_matrix) <- cells
colnames(GCFR_matrix) <- species
for (i in 1:nrow(GCFR_matrix)) {
  GCFR_matrix[i, ] <- species %in% species_occ2$GCFR$species[
    species_occ2$GCFR$QDS == cells[[i]]
  ]
}

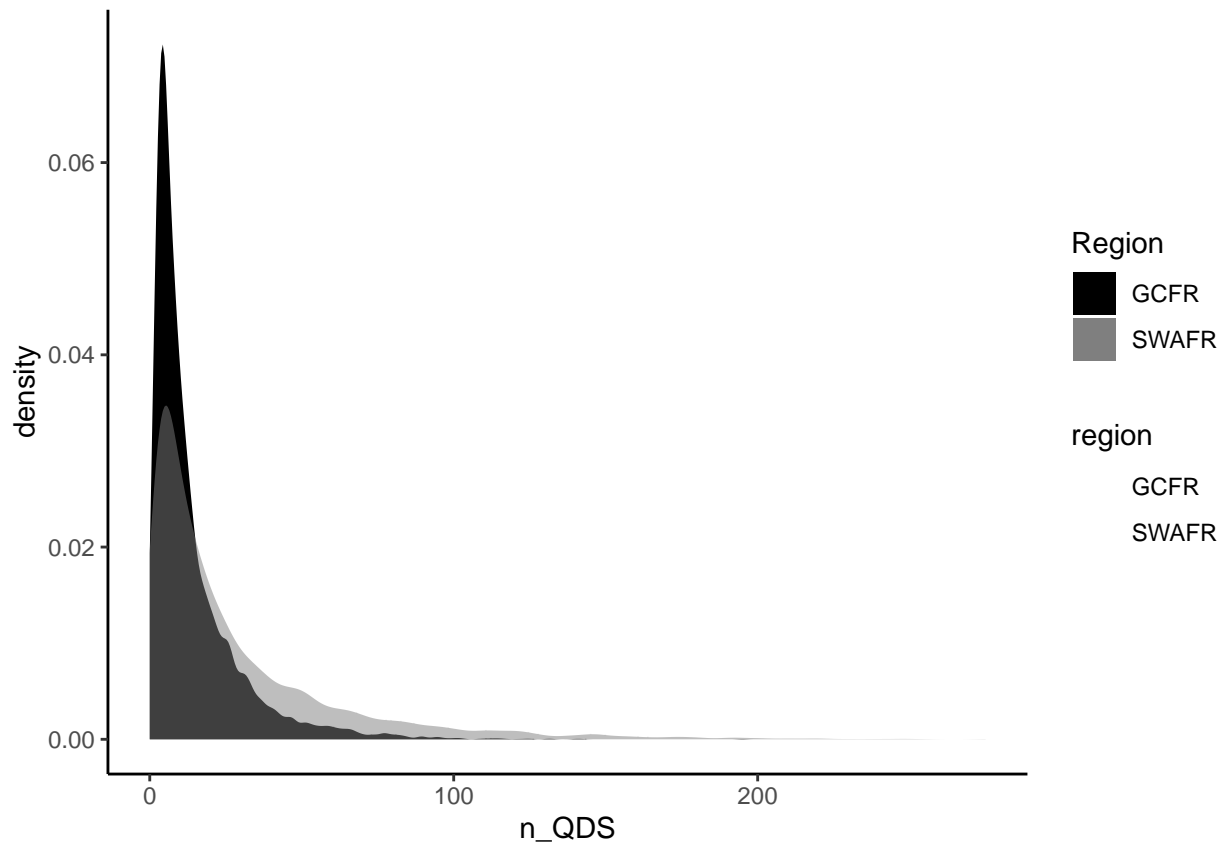
cells <- sort(unique(species_occ2$SWAFR$QDS))
cells <- cells[cells %in% data$QDS]
species <- sort(unique(species_occ2$SWAFR$species))
species <- species[species %in% SWAFR_species$species]
n_cells <- length(cells)
n_species <- length(species)
SWAFR_matrix <- matrix(nrow = n_cells, ncol = n_species)
rownames(SWAFR_matrix) <- cells
colnames(SWAFR_matrix) <- species
for (i in 1:nrow(SWAFR_matrix)) {
  SWAFR_matrix[i, ] <- species %in% species_occ2$SWAFR$species[
    species_occ2$SWAFR$QDS == cells[[i]]
  ]
}

GCFR_range_sizes <- GCFR_matrix %>%
  apply(2, sum) %>%
  sort(decreasing = TRUE) %>%
  {tibble(region = "GCFR", species = names(.), n_QDS = .)}
SWAFR_range_sizes <- SWAFR_matrix %>%
  apply(2, sum) %>%
  sort(decreasing = TRUE) %>%
  {tibble(region = "SWAFR", species = names(.), n_QDS = .)}
range_sizes <-
  full_join(GCFR_range_sizes, SWAFR_range_sizes) %>%
  mutate(log10_n_QDS = log10(n_QDS))

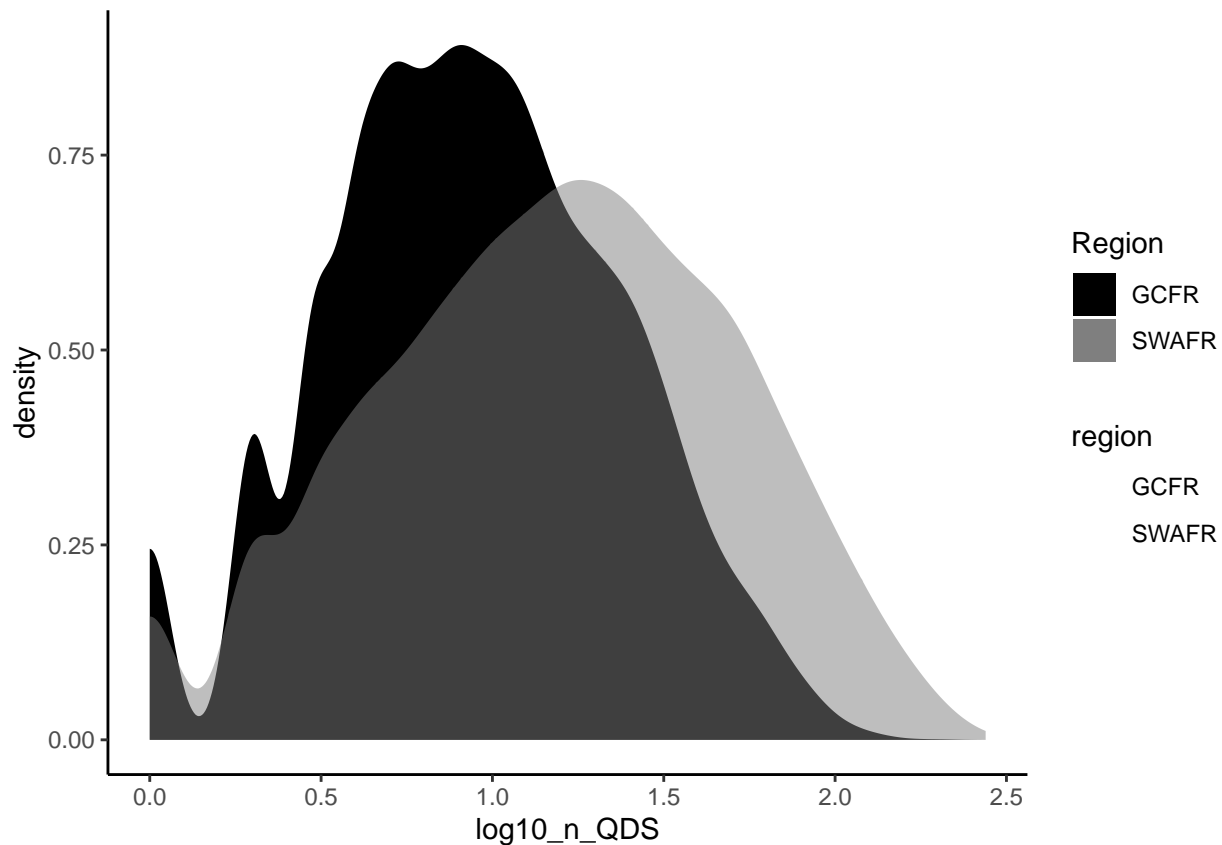
## Joining, by = c("region", "species", "n_QDS")
ggplot(range_sizes) +
  aes(n_QDS, fill = region, alpha = region) +

```

```
geom_density(colour = NA) +
scale_fill_manual(name = "Region", values = c("black", "grey50")) +
scale_alpha_manual(values = c(1, 0.5)) +
theme_classic()
```



```
ggplot(range_sizes[range_sizes$log10_n_QDS >= 0, ]) +
aes(log10_n_QDS, fill = region, alpha = region) +
geom_density(colour = NA) +
scale_fill_manual(name = "Region", values = c("black", "grey50")) +
scale_alpha_manual(values = c(1, 0.5)) +
theme_classic()
```



```
t.test(
  log10_n_QDS ~ region,
  range_sizes[range_sizes$log10_n_QDS >= 0, ]
)
```

```
##
## Welch Two Sample t-test
##
## data: log10_n_QDS by region
## t = -30.579, df = 12721, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2500060 -0.2198856
## sample estimates:
## mean in group GCFR mean in group SWAFR
## 0.9407647 1.1757104
```

```
wilcox.test(
  n_QDS ~ region,
  range_sizes[range_sizes$log10_n_QDS >= 0, ]
)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: n_QDS by region
## W = 23322959, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```