test_south_and_clumping.R

bwien

2023-04-22

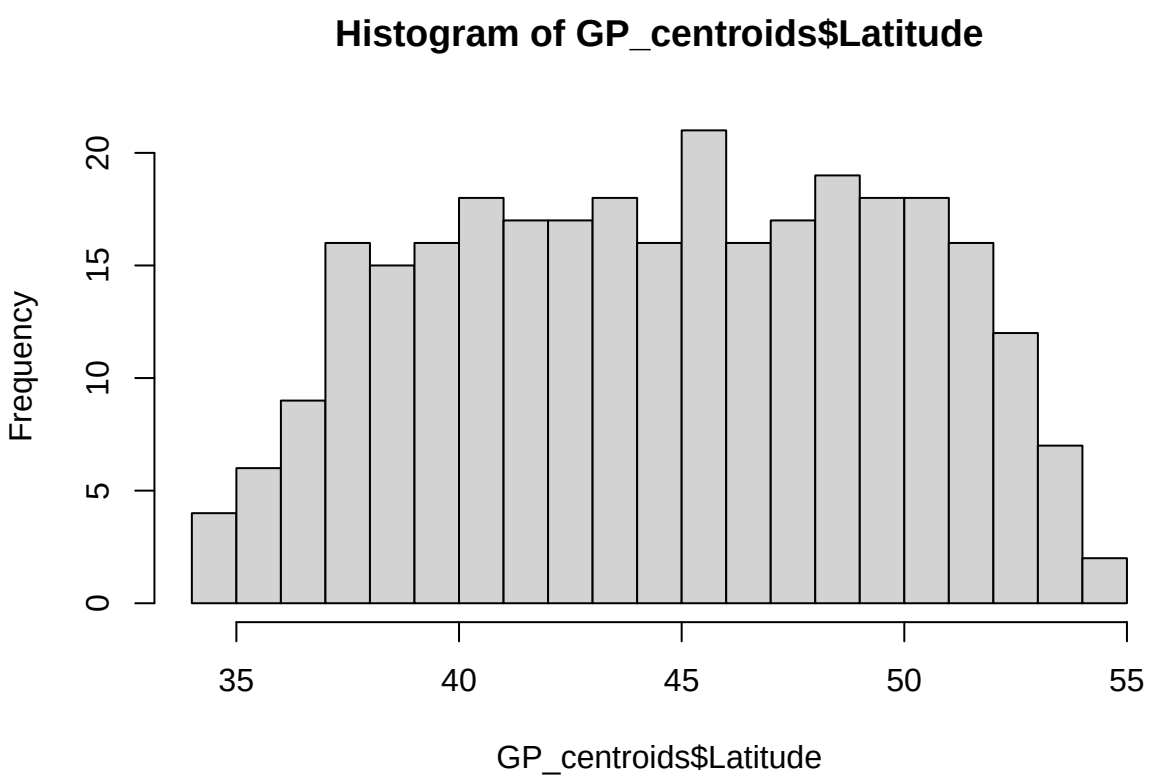
```
setwd("~/Documents/Ben's Stuff/0 KU/Year I/Mustela_nivalis/sampling_effort/FiveDecadesCounted")

library(readr)
```

```
## Warning: package 'readr' was built under R version 4.0.5
```

```
library(geosphere)


# Use only hexes whose centroids are within the Great Plains
# (excludes some significant hexes just on the border)
#GP_centroids <- read_csv("All60sCountsClippedReprojectedCentroidCoordinatesGreatPlainsIntersection.csv

# Use hexes who overlap at all with Great Plains
# (more inclusive of hexes right on the edge of the Great Plains)
GP_centroids <- read_csv("All60sCountsClippedReprojectedHexGreatPlainsIntersection.csv")
```

```
## Rows: 298 Columns: 17
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (5): AllCountsCategories_status60, AllCountsCategories_status60min50, A...
## dbl (12): fid, id, left, top, right, bottom, NivalPre60, BackPre60, NivalPos...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
hist(GP_centroids$Latitude, breaks = 20)
```




```
# Average latitude of centroids with significant range expansion
sig_lat <- GP_centroids[GP_centroids$AllCountsCategories_status60.min76 == "range expansion",]$Latitude
num_sig <- length(sig_lat)
ave_sig_lat <- mean(sig_lat)


#######################################
# Bootstrap sampling to test whether
# average latitude of significant
# cells is farther south than expected
# by chance
#######################################

# Make function to do bootstrapping
weasel_boot <- function (x, length, replicates) {
  ave_boot_lats <- c()
  for (i in 1:replicates) {
    boot_samp <- sample(x$Latitude, length, replace = F)
    ave_boot_samp <- mean(boot_samp)
    ave_boot_lats <- c(ave_boot_lats, ave_boot_samp)
  }
  ave_boot_lats
}

# Do the bootstrapping and look at it
ave_boot_lats <- weasel_boot(GP_centroids, num_sig, 100000)
```

```r
hist(ave_boot_lats)
abline(v=ave_sig_lat)
```


```r
# Look at quantiles for various significance levels
quantile(ave_boot_lats, probs = c(0, 0.01, 0.025, 0.05))
```

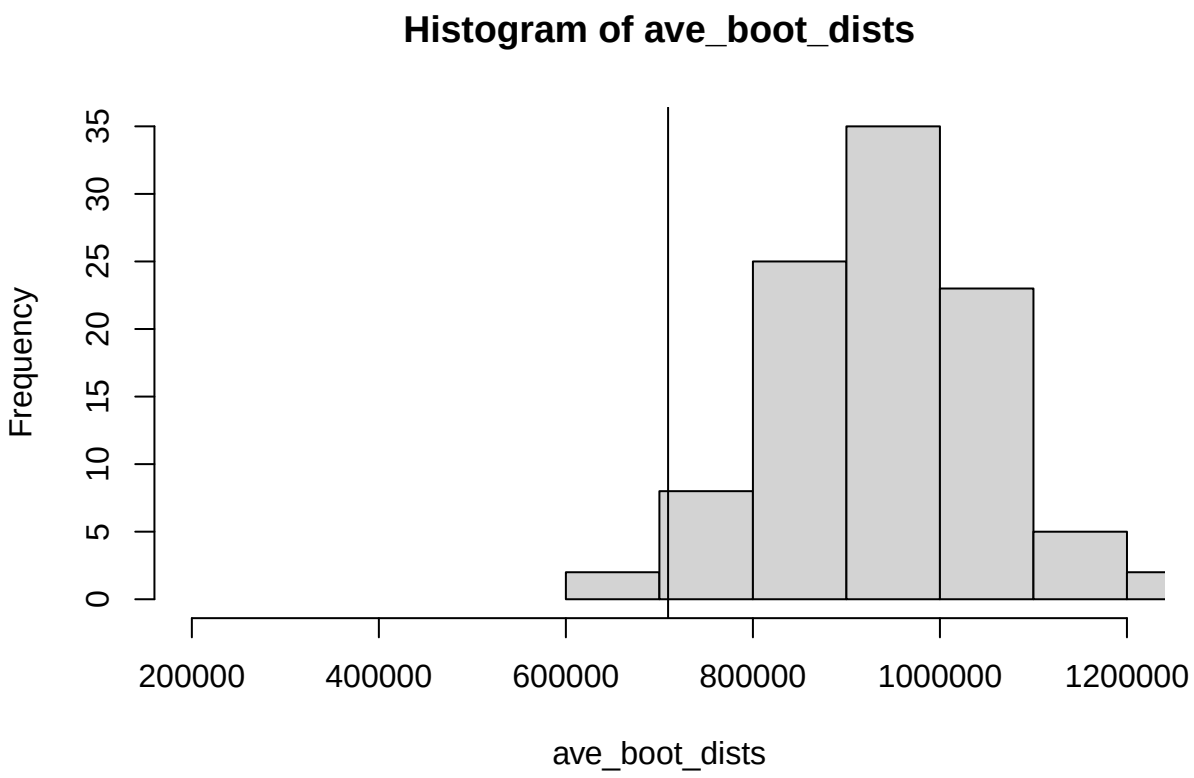
```
##       0%       1%     2.5%       5%
## 38.87454 41.30921 41.83346 42.28251
```

```r
# Get p-value for observed average latitude of significant range expansion cells
percentile <- ecdf(ave_boot_lats)
percentile(ave_sig_lat)
```

```
## [1] 0.0034
```

```r
#######################################
# Bootstrap sampling to test whether
# average distance between significant
# centroids is smaller than expected
# by chance
#######################################
```

```r
# Average pairwise distance of centroids with significant range expansion

# Dataframe of only significant range expansion centroids
GP_centroids_sig <- GP_centroids[GP_centroids$AllCountsCategories_status60.min76 == "range expansion",]


# Make function that calculates average pairwise distance between any set of coordinates
mean_pairwise_dist <- function(x, bootstrap = F, length = NULL, replicates = 1) {
  ave_dist <- c()
  for (r in 1:replicates) {
    if (r%%1000==0) {
      print(r)
    }
    dists <- c()
    y <- x
    if (bootstrap) {
      y <- x[sample(nrow(x), size = length, replace = F),]
    }

    for (i in 1:nrow(y)) {
      j <- i+1
      while (j <= nrow(y)) {
        # centroid points are WGS84, distGeo calculates shortest distance between two points on a WGS84
        # outputs distance in meters
        dist <- distm(c(y$Longitude[i], y$Latitude[i]),
                      c(y$Longitude[j], y$Latitude[j]),
                      fun = distGeo)
        dists <- c(dists, dist)
        j <- j+1
      }
    }
    ave_dist <- c(ave_dist, mean(dists))
    #print(dists)
    #hist(dists)
  }
  return(ave_dist)
}

# To run the function on the real data, plug in GP_centroids_sig, no bootstraps, replicates = 1
ave_sig_dist <- mean_pairwise_dist(GP_centroids_sig, bootstrap = F, replicates = 1)

# To run the funciton on bootstraps, plug in GP_centroids, num_sig, bootstraps = T, replicates)
ave_boot_dists <- mean_pairwise_dist(GP_centroids, bootstrap = T, length = num_sig, replicates = 100)



#Look at bootstrap distribution
hist(ave_boot_dists, xlim = c(200000, 1200000))
abline(v=ave_sig_dist)
```




```
# Look at quantiles for various significance levels
quantile(ave_boot_dists, probs = c(0, 0.01, 0.025, 0.05))
```

```
##       0%       1%     2.5%       5%
## 654922.1 693625.0 711342.2 765999.0
```

```
# Get p-value for observed average latitude of significant range expansion cells
percentile_dists <- ecdf(ave_boot_dists)
percentile_dists(ave_sig_dist)
```

```
## [1] 0.03
```