Occurrences mapping

Norah Saarman

2024-11-10

Point data by time frame

```r
# Load necessary libraries
library(ggplot2)
library(sf)
```

```
## Linking to GEOS 3.10.2, GDAL 3.4.1, PROJ 8.2.1; sf_use_s2() is TRUE
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(maps)

# Load the data from your CSV files, filter out NAs, and ensure `year` column is numeric
pipiens <- read.csv("coords_pip_v5.csv") %>%
  filter(!is.na(year) & latitude > 20) %>%
  mutate(year = as.numeric(year))
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `year = as.numeric(year)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```r
quinquefasciatus <- read.csv("coords_qui_v5.csv") %>%
  filter(!is.na(year) & latitude > 20) %>%
  mutate(year = as.numeric(year))

hybrids <- read.csv("coords_hyb_v5.csv") %>%
  filter(!is.na(year) & latitude > 20) %>%
  mutate(year = as.numeric(year))

# Add time period labels to each dataset and make Time_Period a factor in chronological order
pipiens <- pipiens %>% mutate(Time_Period = factor(case_when(
  year < 1970 ~ "Before 1970",
  year >= 1970 & year <= 2013 ~ "1970-2013",
```

```r
    year >= 2014 & year <= 2024 ~ "2014-2024",
    TRUE ~ "Unknown"   # Ensures no NA values in Time_Period
), levels = c("Before 1970", "1970-2013", "2014-2024")))

quinquefasciatus <- quinquefasciatus %>% mutate(Time_Period = factor(case_when(
    year < 1970 ~ "Before 1970",
    year >= 1970 & year <= 2013 ~ "1970-2013",
    year >= 2014 & year <= 2024 ~ "2014-2024",
    TRUE ~ "Unknown"
), levels = c("Before 1970", "1970-2013", "2014-2024")))

hybrids <- hybrids %>% mutate(Time_Period = factor(case_when(
    year < 1970 ~ "Before 1970",
    year >= 1970 & year <= 2013 ~ "1970-2013",
    year >= 2014 & year <= 2024 ~ "2014-2024",
    TRUE ~ "Unknown"
), levels = c("Before 1970", "1970-2013", "2014-2024")))

# Remove any entries labeled "Unknown" in Time_Period
pipiens <- pipiens %>% filter(Time_Period != "Unknown")
quinquefasciatus <- quinquefasciatus %>% filter(Time_Period != "Unknown")
hybrids <- hybrids %>% filter(Time_Period != "Unknown")

# Combine all datasets
all_data <- bind_rows(
  pipiens %>% mutate(Species = "pipiens"),
  quinquefasciatus %>% mutate(Species = "quinquefasciatus"),
  hybrids %>% mutate(Species = "hybrid")
)

# Load USA state outlines from the maps package
us_states <- map_data("state")

# Plot with facets for each time period, arranged in chronological order
ggplot() +
  # Add USA basemap with state outlines
  geom_polygon(data = us_states, aes(x = long, y = lat, group = group),
               fill = "white", color = "black", alpha = 0.3) +

  # Plot individual points with consistent colors and shapes
  geom_point(data = all_data, aes(x = longitude, y = latitude, color = Species, shape = Species), size =

  # Set color and shape scales
  scale_color_manual(values = c("pipiens" = "blue", "quinquefasciatus" = "red", "hybrid" = "purple")) +
  scale_shape_manual(values = c("pipiens" = 16, "quinquefasciatus" = 17, "hybrid" = 18)) +

  # Facet the plot into three panels by time period in a single row
  facet_wrap(~Time_Period, nrow = 1) +

  # Add title and labels
  labs(title = "Occurrence of Culex Mosquitoes in the United States by Time Period",
       x = "Longitude", y = "Latitude", color = "Species", shape = "Species") +
  coord_fixed(ratio = 1.3) +   # Fixed aspect ratio
```
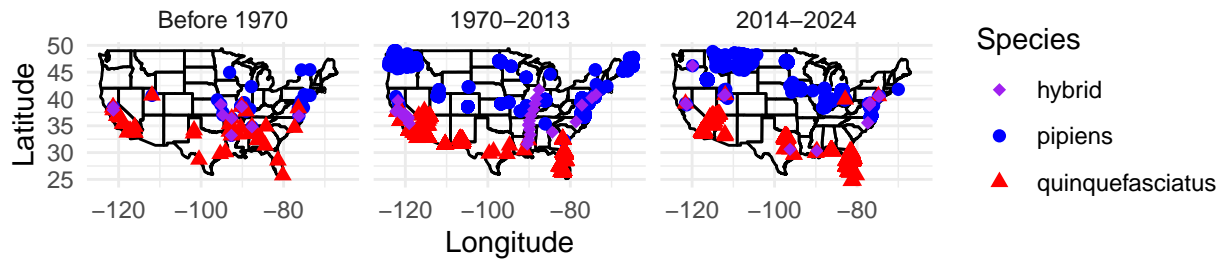
```
# Apply minimal theme
theme_minimal()
```

Occurrence of Culex Mosquitoes in the United States by Time Period

Density plot of hybrids by time period

```
# Plot separate density plots by species
ggplot() +
  # Add USA basemap with state outlines
  geom_polygon(data = us_states, aes(x = long, y = lat, group = group),
               fill = "white", color = "black", alpha = 0.3) +

  # Add density layer for hybrids in purple
  stat_density_2d(data = all_data %>% filter(Species == "hybrid"),
                  aes(x = longitude, y = latitude, fill = after_stat(level)),
                  geom = "polygon", alpha = 0.4) +
  scale_fill_gradient(low = "lavender", high = "purple", name = "Hybrid Density", trans = "log") +

  # Facet the plot into three panels by time period in a single row
  facet_wrap(~Time_Period, nrow = 1) +

  # Add title and labels
  labs(title = "Density of Culex Mosquito Occurrences by Time Period",
       x = "Longitude", y = "Latitude") +
  coord_fixed(ratio = 1.3) +   # Fixed aspect ratio
  # Apply minimal theme
  theme_minimal()
```
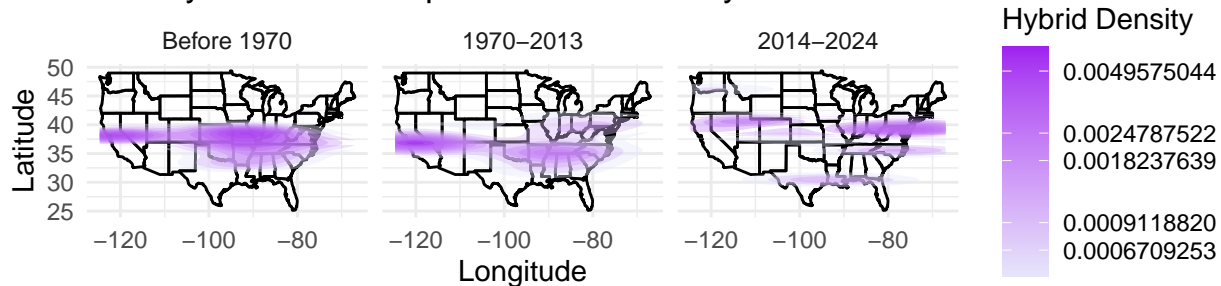
Density of Culex Mosquito Occurrences by Time Period

For each species, aggregating all years

```
library(gridExtra)  # for arranging multiple plots in a grid
```

```
##
## Attaching package: 'gridExtra'
```

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

```r
# Combine all datasets, adding species labels
all_data <- bind_rows(
  pipiens %>% mutate(Species = "pipiens"),
  quinquefasciatus %>% mutate(Species = "quinquefasciatus"),
  hybrids %>% mutate(Species = "hybrid")
)

# Remove any cases where data is insufficient for density calculation
all_data <- all_data %>% filter(!is.na(longitude), !is.na(latitude))

# Load USA state outlines from the maps package
us_states <- map_data("state")

# Create density plot for each species individually with separate color scales and fixed aspect ratio
pipiens_plot <- ggplot() +
  geom_polygon(data = us_states, aes(x = long, y = lat, group = group),
               fill = "white", color = "black", alpha = 0.3) +
  stat_density_2d(data = all_data %>% filter(Species == "pipiens"),
                  aes(x = longitude, y = latitude, fill = after_stat(level)),
                  geom = "polygon", alpha = 0.5) +
  scale_fill_gradient(low = "lightblue", high = "blue", name = "pip", trans = "log") +
  labs(x = NULL, y = NULL) +
  coord_fixed(ratio = 1.3) +  # Fixed aspect ratio
  theme_minimal()

quinquefasciatus_plot <- ggplot() +
  geom_polygon(data = us_states, aes(x = long, y = lat, group = group),
               fill = "white", color = "black", alpha = 0.3) +
  stat_density_2d(data = all_data %>% filter(Species == "quinquefasciatus"),
                  aes(x = longitude, y = latitude, fill = after_stat(level)),
                  geom = "polygon", alpha = 0.5) +
  scale_fill_gradient(low = "pink", high = "red", name = "quinq", trans = "log") +
  labs(x = NULL, y = NULL) +
  coord_fixed(ratio = 1.3) +  # Fixed aspect ratio
  theme_minimal()

hybrid_plot <- ggplot() +
  geom_polygon(data = us_states, aes(x = long, y = lat, group = group),
               fill = "white", color = "black", alpha = 0.3) +
  stat_density_2d(data = all_data %>% filter(Species == "hybrid"),
                  aes(x = longitude, y = latitude, fill = after_stat(level)),
                  geom = "polygon", alpha = 0.5) +
  scale_fill_gradient(low = "lavender", high = "purple", name = "hybrid") +
  labs(x = NULL, y = NULL) +
  coord_fixed(ratio = 1.3) +  # Fixed aspect ratio
  theme_minimal()

# Arrange the plots in a row
grid.arrange(pipiens_plot, quinquefasciatus_plot, hybrid_plot, nrow = 3)
```
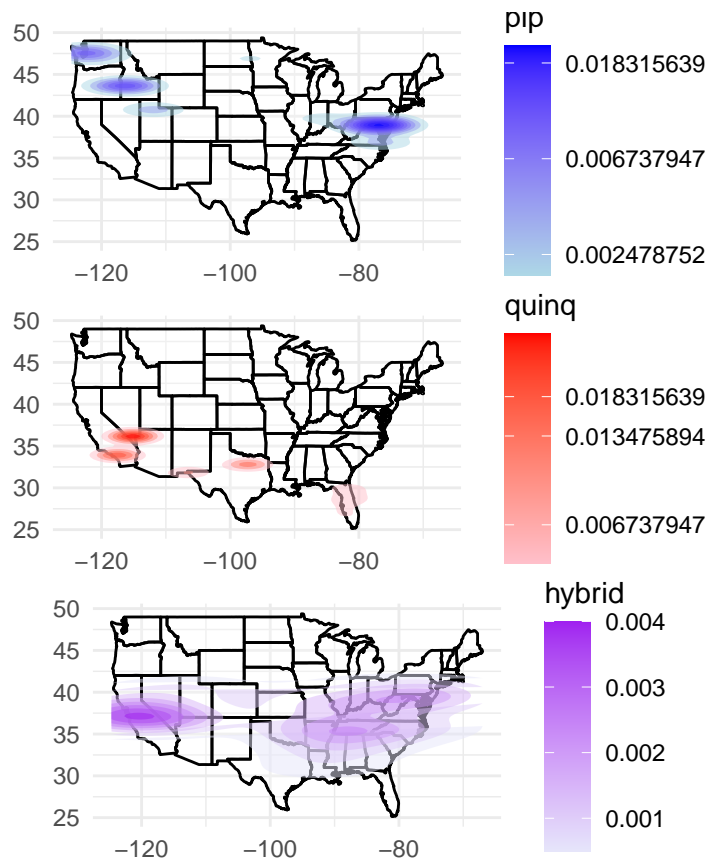
Transparent density plots with thinning

```r
# Load necessary libraries
library(ggplot2)
library(sf)
library(dplyr)
library(maps)


# Function to thin data by 25 km distance
thin_by_distance <- function(data, distance_km = 25) {
  # Convert to sf object with longitude-latitude coordinates
  data_sf <- st_as_sf(data, coords = c("longitude", "latitude"), crs = 4326)

  # Transform to a projected coordinate system for accurate distance calculations
  data_proj <- st_transform(data_sf, crs = 32633)  # Use UTM Zone 33N or an appropriate UTM zone for yo

  # Identify clusters of points within the specified distance
  clusters <- st_is_within_distance(data_proj, dist = distance_km * 1000)  # Convert km to meters

  # Keep only one point per cluster
  unique_indices <- sapply(clusters, function(x) x[1])  # Take the first point in each cluster
  data_thinned <- data[unique_indices, ]

  return(data_thinned)
}
```

```r
# Apply spatial thinning to pipiens and quinquefasciatus data
pipiens_thinned <- thin_by_distance(pipiens, distance_km = 25)
quinquefasciatus_thinned <- thin_by_distance(quinquefasciatus, distance_km = 25)

# Convert state data to sf for more accurate boundary rendering
us_states_sf <- st_as_sf(map("state", plot = FALSE, fill = TRUE))

# Plot with thinned data for density and original data for overlayed points
ggplot() +
  # Base map with more accurate state outlines using sf
  geom_sf(data = us_states_sf, fill = "white", color = "black", alpha = 0.3) +

  # Density layers with large bandwidth for thinned data of each species
  stat_density_2d(data = pipiens_thinned, aes(x = longitude, y = latitude, fill = after_stat(level)),
                  geom = "polygon", alpha = 0.3, fill = "blue", h = c(2, 2)) +
  stat_density_2d(data = quinquefasciatus_thinned, aes(x = longitude, y = latitude, fill = after_stat(le
                  geom = "polygon", alpha = 0.3, fill = "red", h = c(2, 2)) +
  stat_density_2d(data = hybrids, aes(x = longitude, y = latitude, fill = after_stat(level)),
                  geom = "polygon", alpha = 0.3, fill = "purple", h = c(2, 2)) +

  # Overlay individual points from original data for each species
  geom_point(data = pipiens, aes(x = longitude, y = latitude), color = "blue", shape = 16, size = 1, alp
  geom_point(data = quinquefasciatus, aes(x = longitude, y = latitude), color = "red", shape = 17, size
  geom_point(data = hybrids, aes(x = longitude, y = latitude), color = "purple", shape = 18, size = 1, a

  # Titles and coord_sf for accurate geographic alignment
  labs(title = "Culex Mosquito Occurrences with 25 km Thinning",
       x = "Longitude", y = "Latitude") +
  coord_sf() +
  theme_minimal()
```
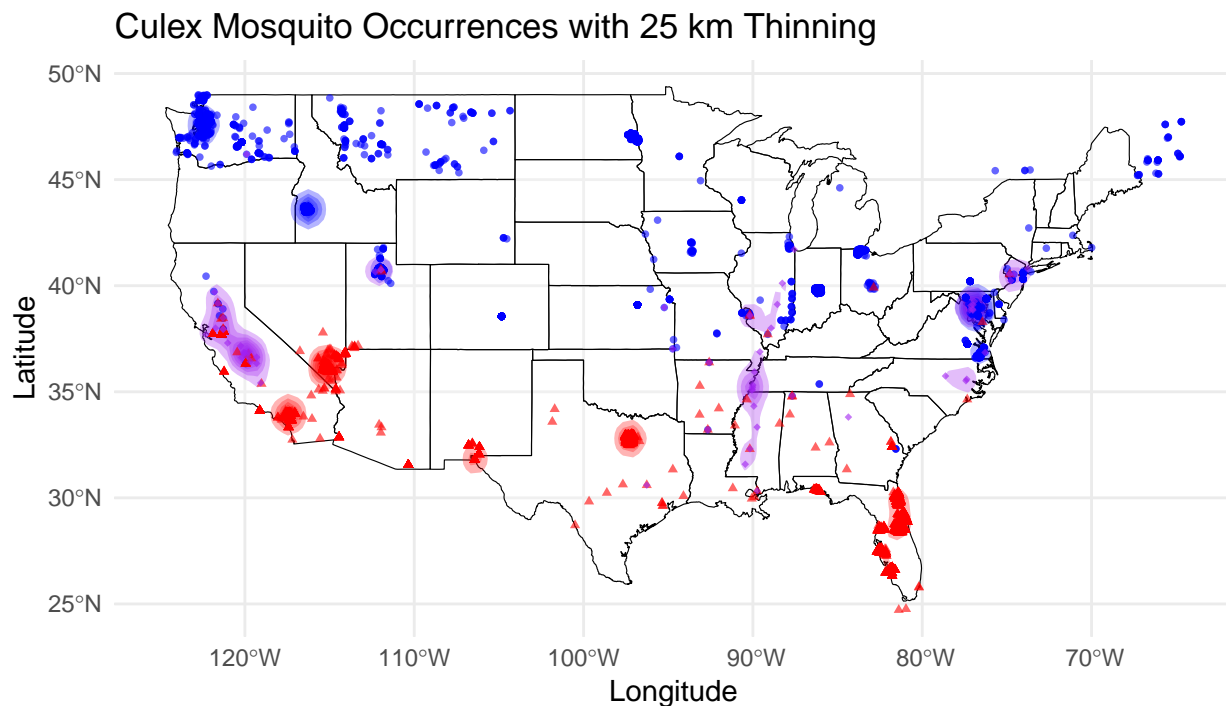

Culex Mosquito Occurrences with 25 km Thinning

Log density of hybrid data points only

```r
# Load necessary libraries
library(ggplot2)
library(sf)
library(dplyr)
library(maps)

# Convert state data to sf for more accurate boundary rendering
us_states_sf <- st_as_sf(map("state", plot = FALSE, fill = TRUE))

# Plot with hybrid density layer and legend for species points
ggplot() +
  # Base map with state outlines using sf
  geom_sf(data = us_states_sf, fill = "white", color = "black", alpha = 0.3) +

  # Logarithmic density layer for hybrids
  stat_density_2d(data = hybrids, aes(x = longitude, y = latitude, fill = after_stat(level)),
                  geom = "polygon", alpha = 1, h = c(2, 2), color = NA) +
  scale_fill_gradientn(colors = c("lavender", "purple"), trans = "log", name = "Density (hybrids)") +

  # Overlay individual points with legend
  geom_point(data = pipiens, aes(x = longitude, y = latitude, color = "pipiens"), shape = 16, size = 1,
  geom_point(data = quinquefasciatus, aes(x = longitude, y = latitude, color = "quinquefasciatus"), sha
  geom_point(data = hybrids, aes(x = longitude, y = latitude, color = "hybrids"), shape = 15, size = 1,

  # Add color legend for species points
  scale_color_manual(values = c("pipiens" = "blue", "quinquefasciatus" = "red", "hybrids" = "purple"),
                     name = "Species") +

  # Titles and standard coord_sf
  labs(title = "Culex Mosquito Occurrences",
       x = "Longitude", y = "Latitude") +
  coord_sf() +
  theme_minimal()
```


Culex Mosquito Occurrences