Comparing Spatial Distributions of the North Atlantic Right Whale and Commercial Shipping Lanes

Diane Egret

August 28th 2022

A. Set up the working environment

Set the working directory. The working directory is the file path R will follow to load or store files.

```
setwd("~/Desktop")
```

Load required packages. A package is a compilation of functions that may be downloaded to execute certain actions. These three packages allow us to process, analyze, manipulate, visualize and model spatial data among other things.

```
library("raster")
library("rgeos")
library("rgdal")
```

We then load environmental data. A background map of the oceans and land will help us illustrate the North Atlantic Right Whale's distribution and put it into context. Here we source it from Gebco.net as a raster of the North Atlantic ocean. We use a tif format (Tagged Image Format), an image format for storing raster graphics images. A raster is two-dimensional map format composed of gridded data that can be visualized when plotted.

```
# load pre-downloaded file of environmental data
# name it "clim"
clim <- raster("gebco_2022_n70.0_s10.0_w-100.0_e20.0.tif")

# print the basic information about "clim"
print(clim)
```

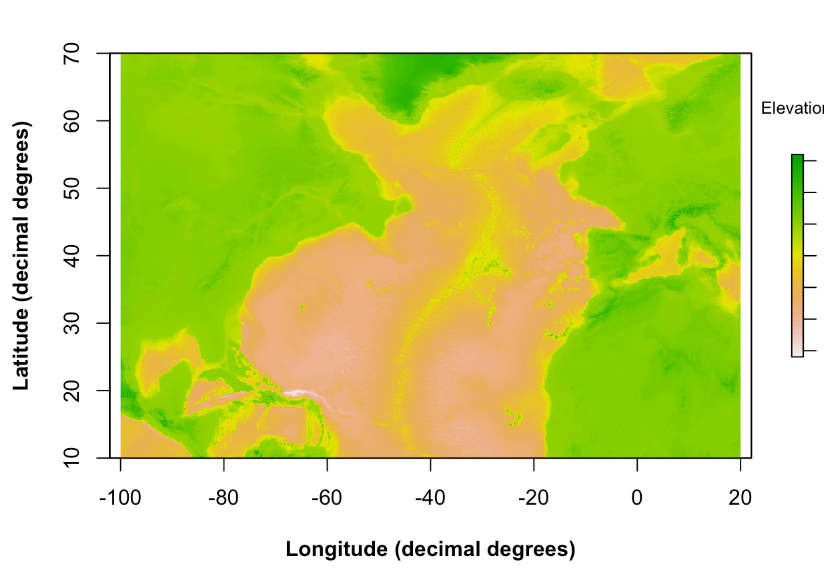
```
## class      : RasterLayer
## dimensions : 14400, 28800, 414720000  (nrow, ncol, ncell)
## resolution : 0.004166667, 0.004166667  (x, y)
## extent     : -100, 20, 10, 70  (xmin, xmax, ymin, ymax)
## crs        : +proj=longlat +datum=WGS84 +no_defs
## source     : gebco_2022_n70.0_s10.0_w-100.0_e20.0.tif
## names      : gebco_2022_n70.0_s10.0_w.100.0_e20.0
## values     : -32768, 32767  (min, max)
```

This information tells us that the crs (Coordinate Reference System) indicates the projection and datum. Here a standard crs is used with longitudes and latitudes. We can also see that the raster has an extent of -90, -25, 20 and 65, or W°-90, E°-25, S°20 and N°65 with the unit being decimal degrees. An extent is the spatial extremes of an object.

But most importantly, we notice that values from the gridded data range from -32768 to 32767 (the unit being meters). This entails that both above-surface and under-surface relief data is included. Let's visualize this:

```
# plot environmental data
plot(clim)

# label the plot
legend(18, 65, "Elevation (meters)", xpd = TRUE, bty = "n", cex = 0.8)
mtext(side = 1, line = 3, "Longitude (decimal degrees)", font = 2, cex = 1)
mtext(side = 2, line = 3, "Latitude (decimal degrees)", font = 2, cex = 1)
```

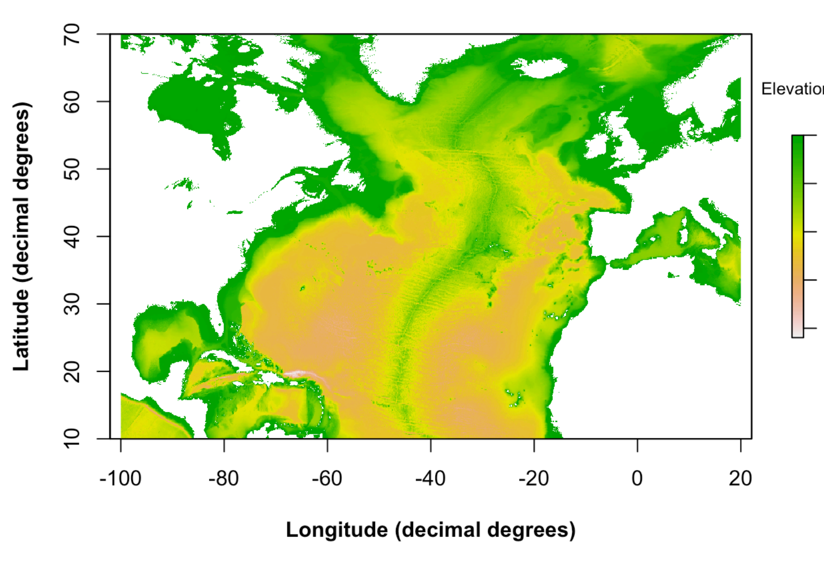


We can indeed observe both land and water elevation. However, we know that *Eubalaena glacialis* is found only in the ocean. We then remove land data which will allow us to better visualize the of the water.

```
# remove data from areas above 0 meters in elevation from "clim"
clim[clim > 0] <- NA

# plot ocean data
plot(clim)

# label the plot
legend(18, 65, "Elevation (meters)", xpd = TRUE, bty = "n", cex = 0.8)
mtext(side = 1, line = 3, "Longitude (decimal degrees)", font = 2, cex = 1)
mtext(side = 2, line = 3, "Latitude (decimal degrees)", font = 2, cex = 1)
```



We can now observe a map of the North Atlantic ocean which includes only ocean data.

B. Plotting the North Atlantic Right Whale's distribution

Download occurrence data. Occurrence data is recorded evidence that a certain species occurred at a certain place at a certain time. Here we download occurrence data of the North Atlantic Right Whale from GBIF.org (Global Biodiversity Information Facility). We use a csv format (comma-separated values), a type of text file.

```
# load pre-downloaded occurrence data
# name it "occ_raw" as in "raw occurrence data"
# Note that the file named "newdata.csv" is simply what I named my occurrence data to simplify its original name
occ_raw <- read.csv("newdata.csv")
```

Remove erroneous data with either the latitude or longitude missing, or duplicated data (based on the coordinates). Create subsets without erroneous data: "occ_clean" then "occ_unique".

```
# remove erroneous coordinates
# name it "occ_clean"
# print amount of occurence data that was removed
occ_clean <- subset(occ_raw, (!is.na(decimalLatitude)) & (!is.na(decimalLongitude)))
cat(nrow(occ_raw) - nrow(occ_clean), "records are removed")
```

```
## 470 records are removed
```

```
# remove duplicates
# name it "occ_unique"
# print amount of occurence data that was removed
dups <- duplicated(occ_clean[c("decimalLatitude", "decimalLongitude")])
occ_unique <- occ_clean[!dups, ]
cat(nrow(occ_clean) - nrow(occ_unique), "records are removed")
```

```
## 410 records are removed
```

Up until now, we were working with a data frame, but this format has no spatial relationship with environmental layers. So we transform our data frame into spatial data.

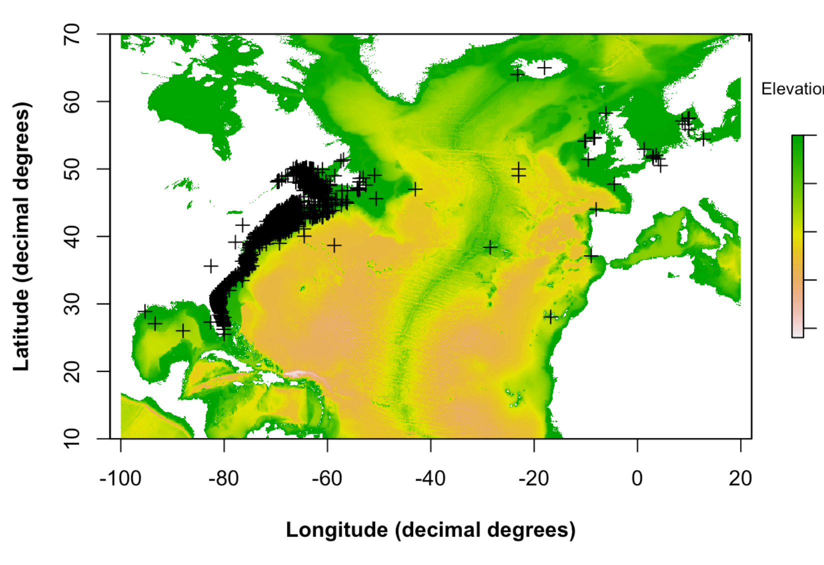
```
# this tells R what the names of the longitude and latitude columns in our data frame so that it can associate it as coordinates
coordinates(occ_unique) <- ~decimalLongitude + decimalLatitude
```

Plot data to look for erroneous occurrences located out of the known distribution of *Eubalaena Glacialis*.

```
# plot ocean data
plot(clim)

# plot occurrence data on top of the ocean data
plot(occ_unique, add = TRUE)

# label the plot
legend(18, 65, "Elevation (meters)", xpd = TRUE, bty = "n", cex = 0.8)
mtext(side = 1, line = 3, "Longitude (decimal degrees)", font = 2, cex = 1)
mtext(side = 2, line = 3, "Latitude (decimal degrees)", font = 2, cex = 1)
```



Note that here and for the rest of this study, each point in the shape of a "plus sign" (+) represents one occurrence of a North Atlantic Right Whale. Each point is not an individual whale but rather one of multiple locations where a whale was detected.

Considering we know that *Eubalaena Glacialis* is distributed throughout the North Atlantic ocean, we can identify a couple points located outside of the known distribution, notably in the Gulf of Mexico. Although *Eubalaena Glacialis* can be found in some parts of Europe, its presence is much more important in the Western Atlantic. For our purposes, being the study of the distribution of the species along the eastern coasts of the United States and Canada, we will only keep occurrence data located in the eastern North Atlantic Ocean. To remove the rest of the points from our occurrence dataset, we only keep points with longitudes between -85° and -45°.

```
# only keep points with longitudes between -85° and -45°
occ_unique <- occ_unique[which(occ_unique$decimalLongitude > -85 & occ_unique$decimalLongitude < -45), ]
```

Thin occurrence data to keep only one occurrence point per pixel.

```
# keep one occurrence point per cell
# name final subset with thinned occurence data "occ_final"
cells <- cellFromXY(clim, occ_unique)
dups <- duplicated(cells)
occ_final <- occ_unique[!dups, ]
cat(nrow(occ_unique) - nrow(occ_final), "records are removed")
```

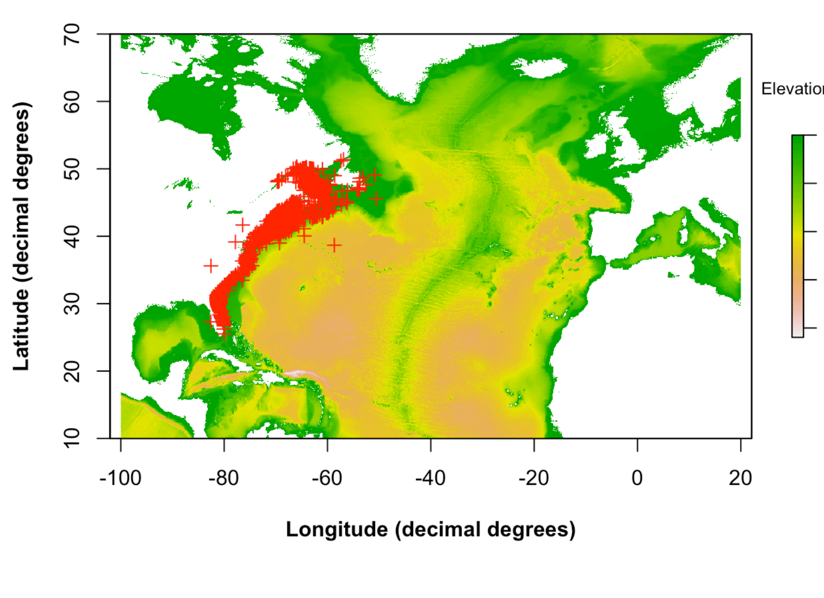
```
## 571 records are removed
```

After sorting our data, we now have 6469 occurrences. Let's plot them:

```
# plot ocean data
plot(clim)

# plot the final occurrence data on top of the ocean data
plot(occ_final, add = TRUE, col = "red")

# label the plot
legend(18, 65, "Elevation (meters)", xpd = TRUE, bty = "n", cex = 0.8)
mtext(side = 1, line = 3, "Longitude (decimal degrees)", font = 2, cex = 1)
mtext(side = 2, line = 3, "Latitude (decimal degrees)", font = 2, cex = 1)
```



C. Plotting the shipping traffic density

We then load our ship traffic data. Here we source it from NCEAS (National Center for Ecological Analysis and Synthesis) from an impressive study on the assessment of cumulative human impacts globally. Download the Commercial Shipping Lanes dataset. Here we use the one from 2013, with a tif format.

```
# load pre-downloaded file of ship traffic data
# name it "shippingmap"
shippingmap <- raster("raw_2013_shipping_mol/shipping.tif")

# print the basic information about "shippingmap"
print(shippingmap)
```

```
## class      : RasterLayer
## dimensions : 19305, 38610, 745366050  (nrow, ncol, ncell)
## resolution : 934.4789, 934.4789  (x, y)
## extent     : -18040095, 18040134, -9020067, 9020047  (xmin, xmax, ymin, ymax)
## crs        : +proj=moll +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs
## source     : shipping.tif
## names      : shipping
## values     : 0, 45226.26  (min, max)
```

We see that the is made of very large numbers (-18040095 and 18040134 for xmin and xmax for example). This is because this projection is in meters. We will convert this projection to match the one used in part 1 with longitudes and latitudes in decimal degrees.

```
# assign the ship traffic data the same crs as our ocean and occurrence data: +proj=longlat +datum=WGS84.
# name it "shipping_longlat"
shipping_longlat <- projectRaster(shippingmap, crs='+proj=longlat +datum=WGS84')
```

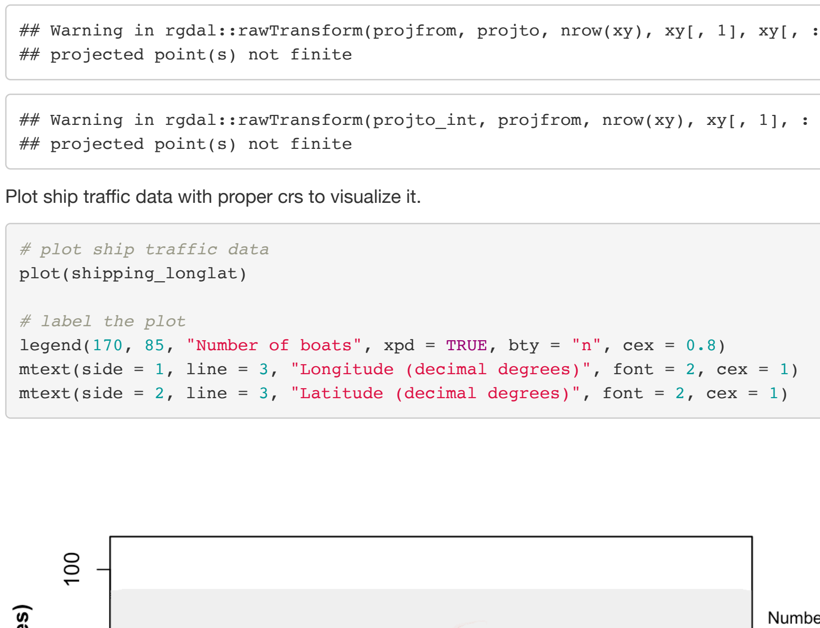
```
## Warning in rgdal::rawTransform(projfrom, projto, nrow(xy), xy[, 1], xy[, : 302
## projected point(s) not finite
```

```
## Warning in rgdal::rawTransform(projto_int, projfrom, nrow(xy), xy[, 1], : 230838
## projected point(s) not finite
```

Plot ship traffic data with proper crs to visualize it.

```
# plot ship traffic data
plot(shipping_longlat)

# label the plot
legend(170, 85, "Number of boats", xpd = TRUE, bty = "n", cex = 0.8)
mtext(side = 1, line = 3, "Longitude (decimal degrees)", font = 2, cex = 1)
mtext(side = 2, line = 3, "Latitude (decimal degrees)", font = 2, cex = 1)
```



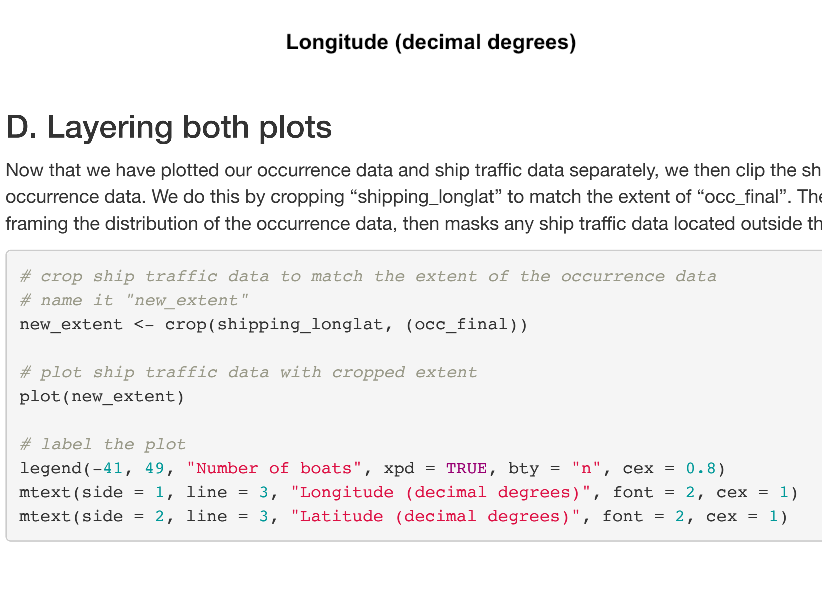
D. Layering both plots

Now that we have plotted our occurrence data and ship traffic data separately, we then clip the ship traffic data to match the extent of our occurrence data. We do this by cropping "shipping_longlat" to match the extent of "occ_final". The "crop" function creates a rectangular shape framing the distribution of the occurrence data, then masks any ship traffic data located outside the rectangle.

```
# crop ship traffic data to match the extent of the occurrence data
# name it "new_extent"
new_extent <- crop(shipping_longlat, (occ_final))

# plot ship traffic data with cropped extent
plot(new_extent)

# label the plot
legend(-41, 49, "Number of boats", xpd = TRUE, bty = "n", cex = 0.8)
mtext(side = 1, line = 3, "Longitude (decimal degrees)", font = 2, cex = 1)
mtext(side = 2, line = 3, "Latitude (decimal degrees)", font = 2, cex = 1)
```

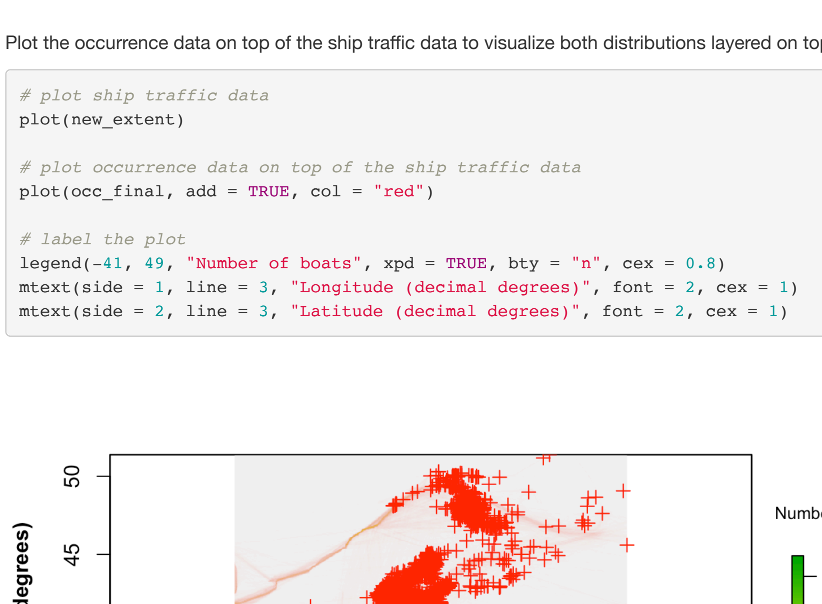


Plot the occurrence data on top of the ship traffic data to visualize both distributions layered on top of each other.

```
# plot ship traffic data
plot(new_extent)

# plot occurrence data on top of the ship traffic data
plot(occ_final, add = TRUE, col = "red")

# label the plot
legend(-41, 49, "Number of boats", xpd = TRUE, bty = "n", cex = 0.8)
mtext(side = 1, line = 3, "Longitude (decimal degrees)", font = 2, cex = 1)
mtext(side = 2, line = 3, "Latitude (decimal degrees)", font = 2, cex = 1)
```



To better illustrate whether both distributions overlap, reduce the size and opacity of the points representing individual ships.

```
# plot ship traffic data
plot(new_extent)

# plot occurrence data with reduced size and opacity on top of the ship traffic data
plot(occ_final, add = TRUE, col = rgb(red = 1, green = 0, blue = 0, alpha = 0.3), cex = 0.5)

# label the plot
legend(-41, 49, "Number of boats", xpd = TRUE, bty = "n", cex = 0.8)
mtext(side = 1, line = 3, "Longitude (decimal degrees)", font = 2, cex = 1)
mtext(side = 2, line = 3, "Latitude (decimal degrees)", font = 2, cex = 1)
```