



Tutorial: Mapping with R

Preparing geographical data for modelling

SÉBASTIEN ROCHETTE, THINKR

ThinkR

Contents

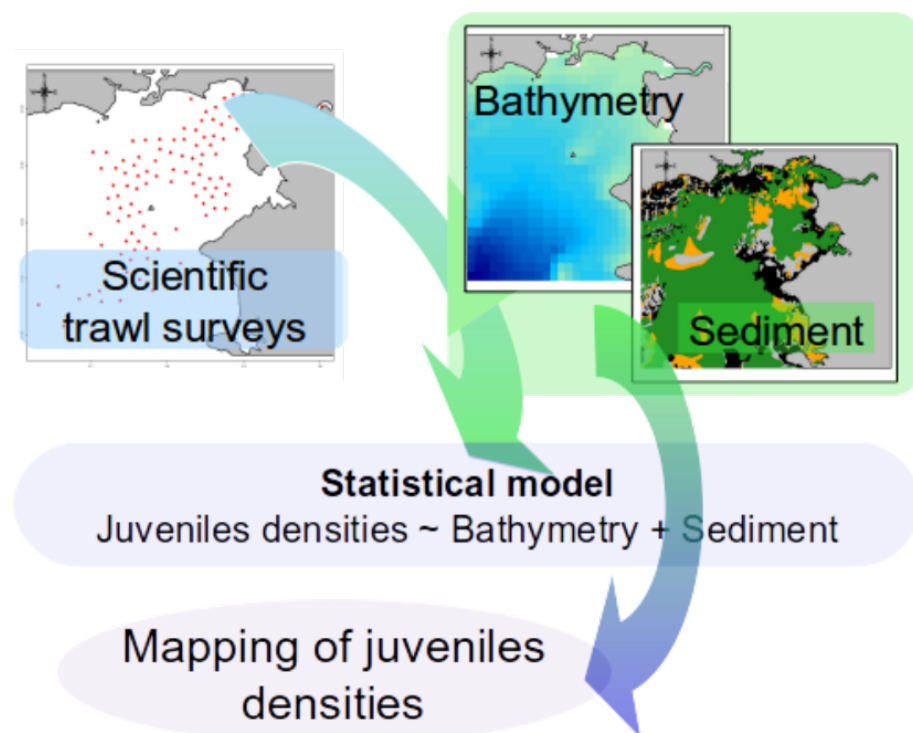
1 Objectives	1
2 Part 1: Manipulate geographic data	1
2.1 Load packages	1
2.2 Read map of France with departments	2
2.3 Read sampling dataset	2
2.4 Read raw covariates and adjust to the study area	3
2.5 Create extra covariate	4
2.6 Extract covariates information at sampling positions	5
2.7 Create a complete multi-layer raster	6

1. Objectives

The aim of this tutorial is to predict the distribution of flatfishes in the Vilaine Estuary (France).

With this tutorial, you will have to (1) manipulate geographic data to prepare data necessary for the modelling procedure, (2) model average probability of presence of flatfishes on the study area and (3) produce maps of predictions of probabilities.

The 3 parts may be independant if needed.



2. Part 1: Manipulate geographic data

2.1. Load packages

Load necessary packages like `{sf}`, `{raster}`, `{dplyr}`, `{readr}`, `{ggplot2}`, `{rasterVis}`

This part aims at preparing a dataset for future modeling purpose. The future model will statistically link a variable of interest (a fish species presence) to environmental covariates. Then, the model can be used for predictions at the scale of the study area. This requires to retrieve covariates for the total area, to retrieve point observation sampled on the area and to know covariates at samples position.

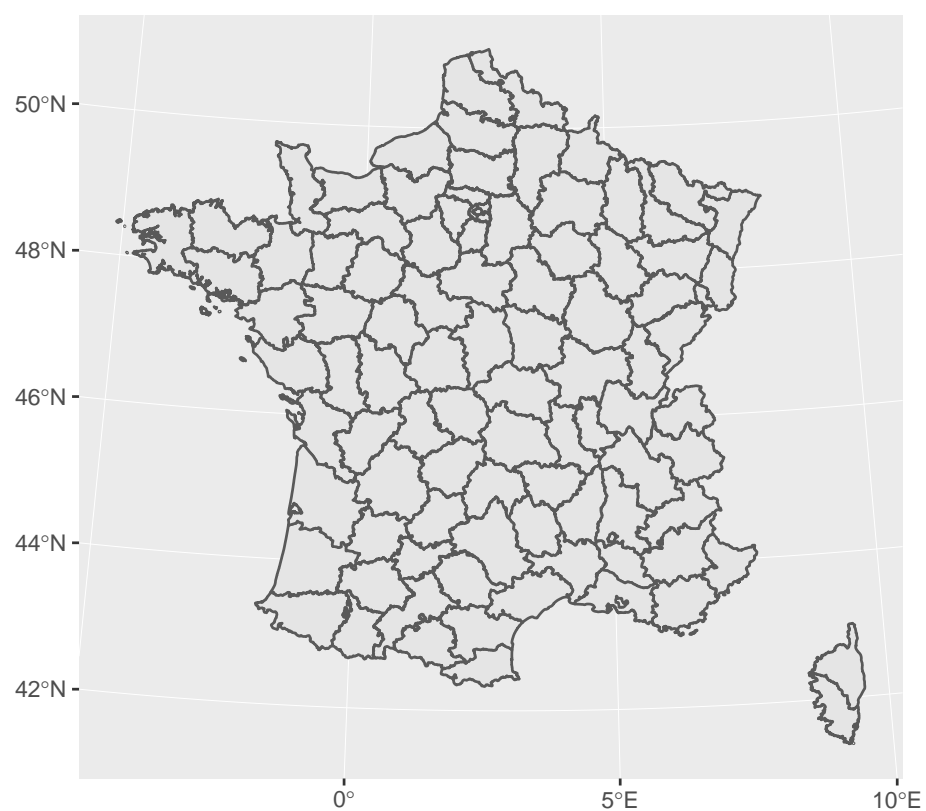
Covariates are:

- Sediment structure. Is the bottom made of mud, sand, gravels or rocks
- Bathymetry. Similar to water depth but calculated from a reference line. Bathymetry does not change with modification of sea level because of tides.
- Biological areas. Areas delimited by expertise considering combination of bathymetry and sediment structure. Related to fishes preys habitat and availability.
- Area close to the coasts. Human disturbances on lands may have an impact on fishes. Distance to the coasts is one of the proxies to test for this.

2.2. Read map of France with departments

To be able to create nice output maps, use map of France as background.

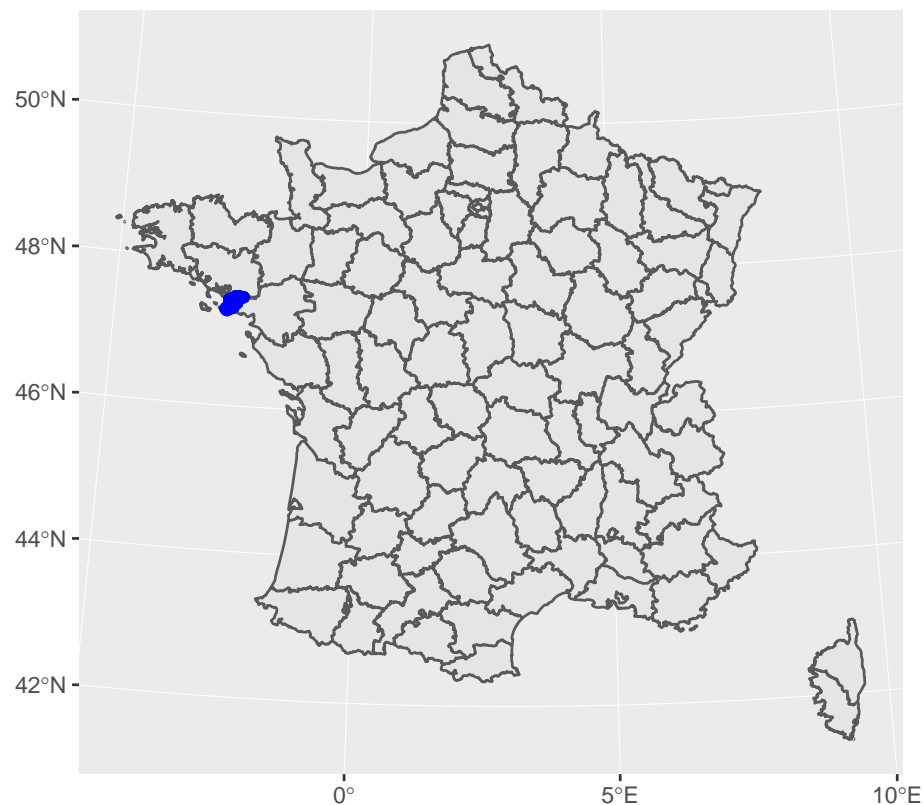
- Read “carto-data-orig/departements/DEPARTEMENT.shp”
 - Store it in the R session with name `dept_193`



2.3. Read sampling dataset

The dataset is saved as a csv file:

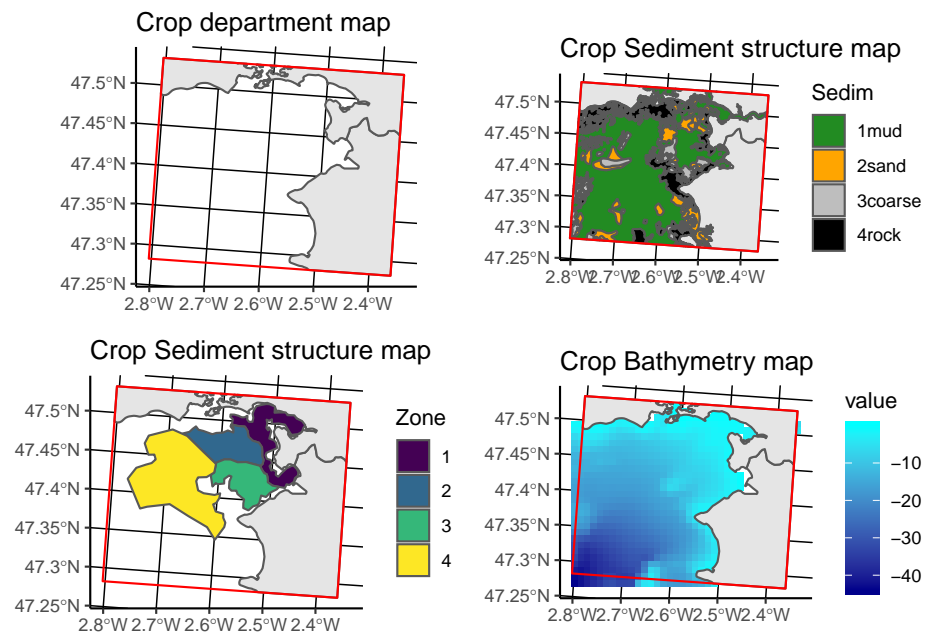
- Data directory: “carto-data-orig/vilaine/Data_Vilaine_solea.csv”
- Transform it as a spatial dataframe. Check function `st_as_sf`. Do not forget parameter `crs`.
- Create a new column named “Presence” that stores 0 when Density is zero and 1 when Density observed is positive



2.4. Read raw covariates and adjust to the study area

- Use the study area polygon as the extent to crop covariates spatial files
 - Data directory: “carto-data-orig/vilaine/ExtAreaPol_wgs84.shp”
- Get original spatial data of sediments, bathymetry and biological areas
 - Data directory: “carto-data-orig/vilaine/Sedim_GDG_wgs84.shp”
 - Data directory: “carto-data-orig/vilaine/bathy_GDG_1000_merc.envi”
 - Data directory: “carto-data-orig/vilaine/Zone_bio_wgs84.shp”
 - Transform the “Zone” covariate into a “factor” class
- Crop layers to adjust to the study area extent.
 - Sediments and biological areas are shapefiles. Think about spatial intersection.
 - Bathymetry is a raster. Think about cropping.
- Crop department spatial layer to restrict the field of view of maps

Be careful with projections... Use appropriate names for spatial objects in the R session.

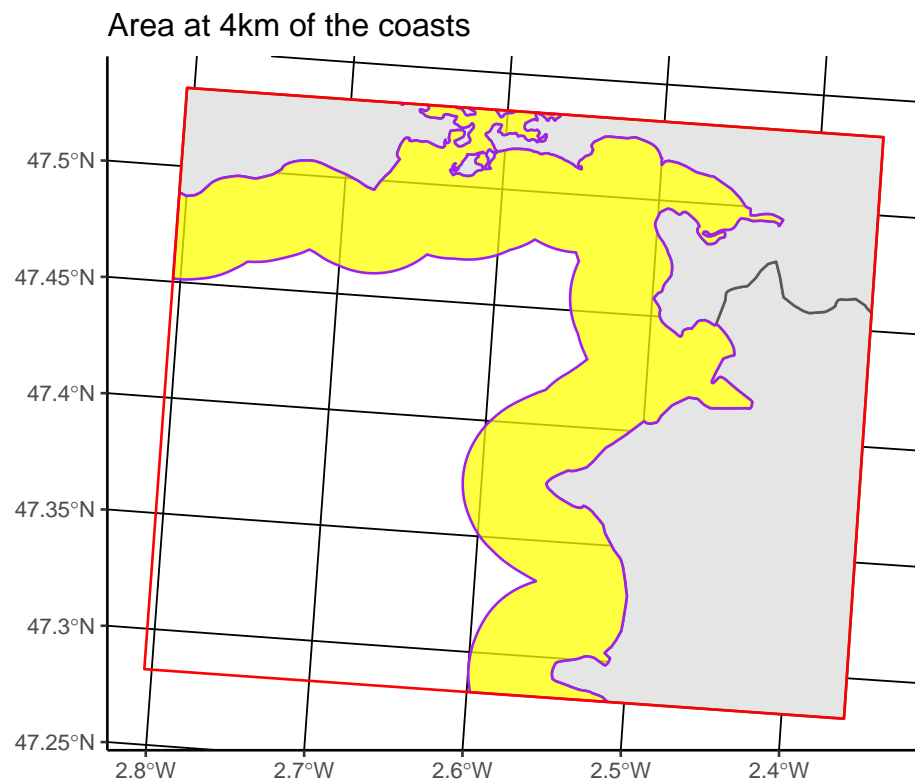


2.5. Create extra covariate

2.5.1 Close to the coasts

We need to find if human disturbances on the continent, simplified as distance to the coast, is a parameter influencing fish distribution.

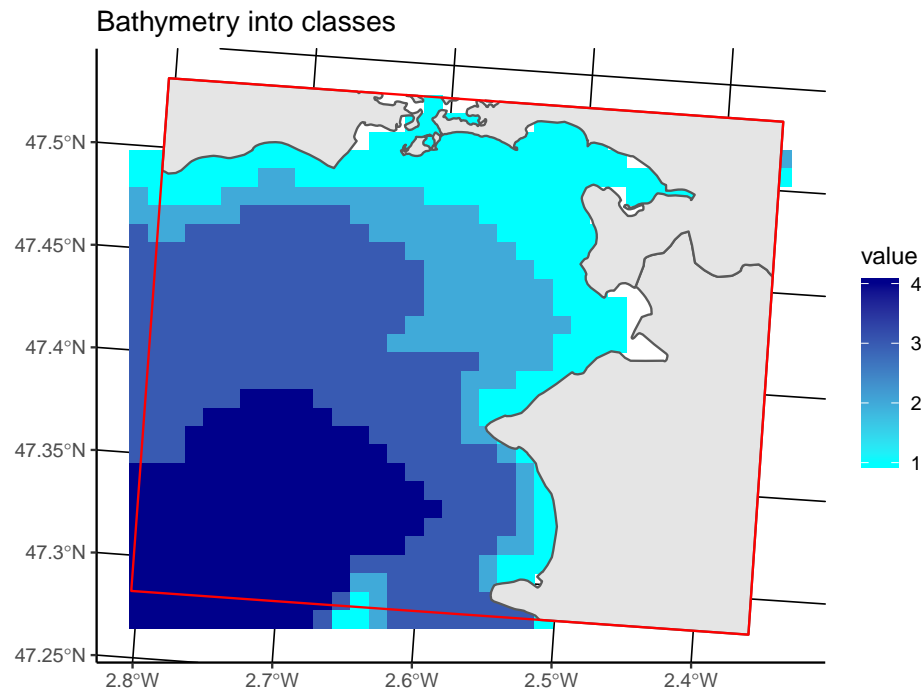
- Create a polygon that covers the area between 0 and 4km to the coastline
 - Merge all polygons of France department
 - Create the correct area
 - Crop to the study area



2.5.2 Bathymetry as class factor

Depth may be a proxy for multiple environmental parameters. Distance to the river mouth, hiding areas, ... Dividing bathymetry into classes may makes more sense. For data exploration, it is interesting too.

- Create a new raster covariate, which is the (opposite) bathymetry divided into classes
 - Breaks to use on $-1*bathy$: `c(-10,5,10,20,50,100)`



2.6. Extract covariates information at sampling positions

The model will need all covariates information to be linked to presence data observation.

- Extract information from each covariate at sampling positions
 - Sediment, Bathymetry, Bathymetry classes, Biological areas, Area of the coasts
 - Use `st_join` as some points may be out of the shapefiles
 - Biological areas: Set value to 0 if a station is not over an area
 - Coasts: Set value to `offshore` if not inside the area
 - Bathymetry: Rename column to `bathy`
 - Bathymetry classes: Rename column to `bathyclass`
- Remove unnecessary columns
- Save the resulting spatial layer on your drive as ESRI shapefile (default)
 - Use file name: `"data_vilaine_covariates_l93.shp"`
- Try to reproduce the multipanel figure



2.7. Create a complete multi-layer raster

To do predictions after the modelling procedure, we will use a RasterStack containing all covariates in different layers.

- Use the Bathymetry raster as a basis for all layers
- Use `rasterize` to transform polygons as raster layers.
- Use `stack` to stack all rasters together
- Define layer names identical to covariates in stations data
- Save the resulting stack on your drive using R `.grd` format (some formats do not allow to keep layer names)

