



Tutorial: Modelling in R

Modelling presence-absence

SÉBASTIEN ROCHETTE, THINKR

ThinkR

Contents

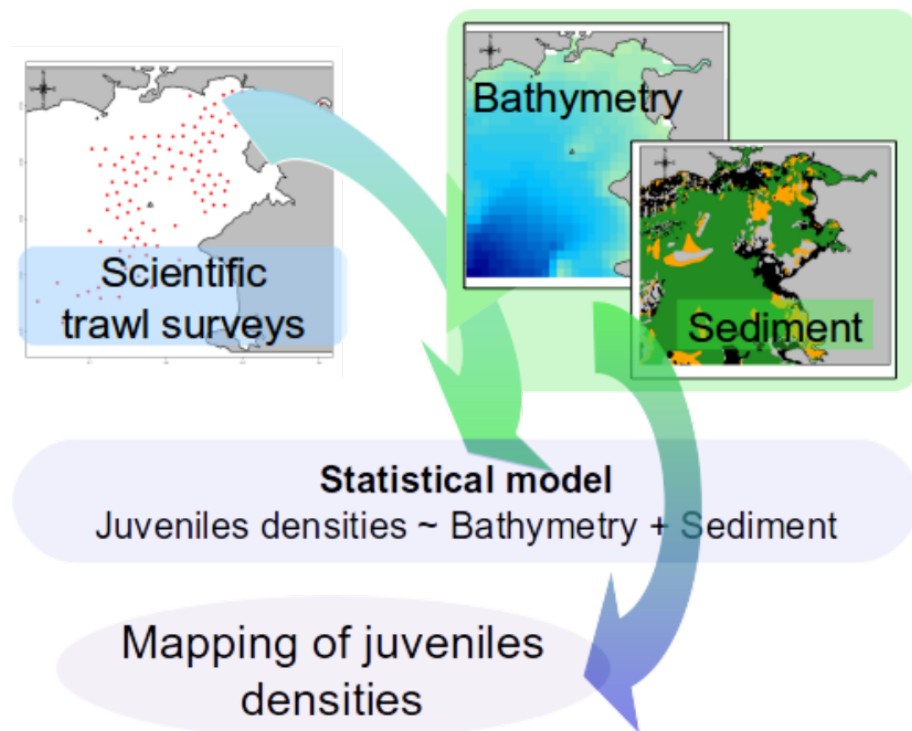
1 Objectives	1
2 Part 2: Modelling probability of presence	1
2.1 Load packages	1
2.2 Outlines	1
2.3 Exploration	2
2.4 Fit a binomial model	2
2.5 Analyse output of the best model	3

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 2: Modelling probability of presence

2.1. Load packages

Load necessary packages like {sf}, {dplyr}, {tidyr}, {ggplot2}, {skimr}, {MASS}, {pROC}

2.2. Outlines

- Explore the dataset
- Use a binomial distribution
 - Test for covariates, quality criterion
- Choose the best model to predict presence of flatfishes

- Explore model outputs
- Look at predictions

2.3. Exploration

- Explore graphically the distribution of each covariate, relations between covariates and sampling plan balance
- Try to guess which covariates may be important to explain the presence of fishes

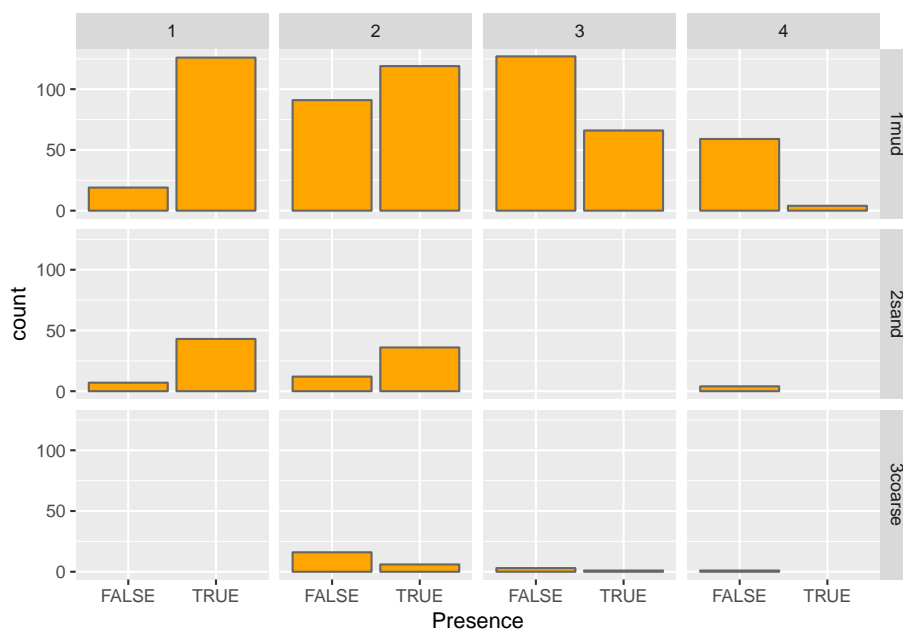
2.3.1 Read dataset

- Read spatial dataset of stations with covariates:
 - “carto-data-orig/vilaine/data_vilaine_covariates_correction_l93.shp”
- Transform as simple tibble for model exploration with `as_tibble`
- Transform columns `Presence` as logical (TRUE/FALSE)
- Transform columns `Year`, `Sedim`, `Zone`, `coasts` and `bathyclass` as character
- Store the result of these operations in an object named `dataset`

2.3.2 Use the 5-functions-you-have-to-always-run on the dataset object

2.3.3 Explore the balance of the sampling plan

- Calculate the number of stations for each group of bathyclass and Sedim
 - Show the output as a wide format with `spread` for a better interpretation
 - What do you think about a possible model that could include both these parameters ?
- Calculate the repartition of presence-absence against *character* covariates
 - Show tables and figures
 - Can you guess what may be important covariates to predict distribution ?



2.4. Fit a binomial model

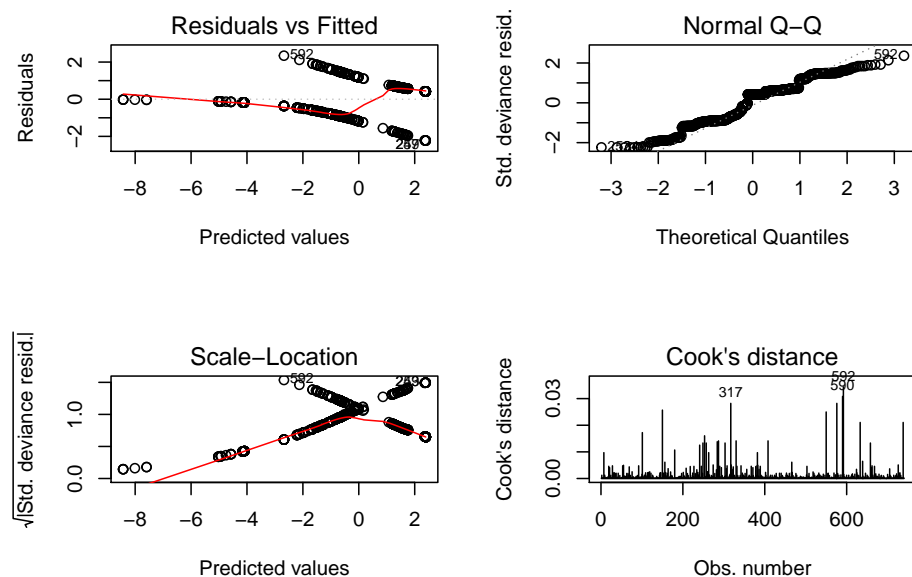
The choice of a distribution for presence-absence data is simple, this is a binomial distribution.

- Find the best model (combination of covariates) to predict presence-absence according to AIC

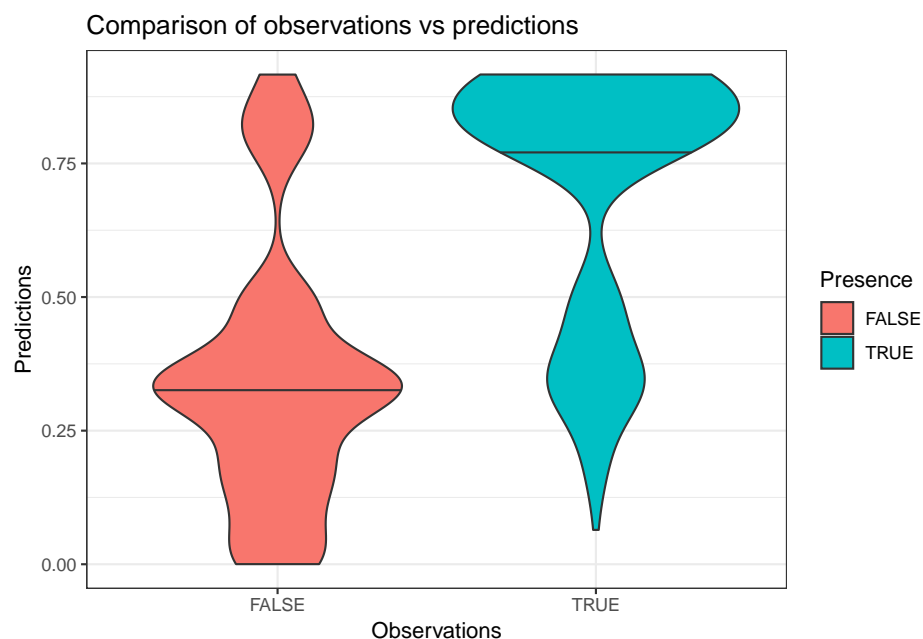
Remember the original question of this case study: We try to predict the average distribution of fishes according to environmental covariates. Choose your model accordingly.

2.5. Analyse output of the best model

- Analysis of residuals of a binomial model is also to be done, even if there is no really choice in the distribution. Visual outputs of residuals analysis of a binomial model are specific .



- Compare predictions and observations
 - What would be the best probability of presence to separate presence from absence ?



2.5.1 Predictions on external dataset

- Read simplified dataset "carto-data-orig/vilaine/predictions.csv"
- Transform covariates with the correct type
- Select only columns with covariates retained in the model
- Remove duplicates
- Add a new column with calculated predictions
- Plot the averaged predictions considering each covariate retained in the model

