



Data Science

Migration and Climate Change

A public policy analysis

Mariana Monroy Paola Hidalgo





“Climate-related displacement is neither a future phenomenon either a problem to be solved, in fact it is a human reality that has to be managed”

Marine Franck, United Nations High
Commissioner

What variables make a country take action in the area of migration and displacement caused by climate change?

World Bank and NDC

DATA

Qualitative and
quantitative

How is the data?

Web Scraping Climate Watch

We webscraping and transform the url of NDC to actual data. We estimated how many pages per country there are and based on that we program a lapply to perform a rbind on all the tables. Later, a csv of each country were saved.



The screenshot shows the 'Data Explorer' section of the Climate Watch website. At the top, there are navigation links: CLIMATEWATCH, COUNTRIES, SECTORS, COMMITMENTS, GHG EMISSIONS, PATHWAYS, ABOUT, DOWNLOAD & VISUALIZE, and a bell icon. Below this is a search bar with placeholder text 'Search for data...'. Under the search bar, there's a 'Data Explorer' title and a sub-section titled 'Raw Data Sources & Metadata'. This section includes dropdown menus for 'Country', 'Global category', 'Overview category', 'Sector', 'Subsector', and 'Value'. A table below lists data for Afghanistan, with columns for 'Country', 'Global category', 'Overview category', 'Sector', 'Subsector', and 'Value'. The 'Value' column contains entries like 'INDC Submitted', '11/23/2016', and 'Only First NDC'. At the bottom of this section is a URL: <https://www.climatewatchdata.org/data-explorer/ndc-content>. Above this table is a large list of CSV files for various countries: AFG_ndc.csv, AGO_ndc.csv, ALB_ndc.csv, AND_ndc.csv, ARE_ndc.csv, ARG_ndc.csv, ARM_ndc.csv, ATG_ndc.csv, AUS_ndc.csv, AUT_ndc.csv, AZE_ndc.csv, BDI_ndc.csv, BEL_ndc.csv, BEN_ndc.csv, BFA_ndc.csv, BGD_ndc.csv, BGR_ndc.csv, BHR_ndc.csv, BHS_ndc.csv, BIH_ndc.csv.

The screenshot shows a file manager interface displaying a grid of CSV files for 197 countries. The files are organized in several rows: AFG_ndc.csv, AGO_ndc.csv, ALB_ndc.csv, AND_ndc.csv, ARE_ndc.csv; ARG_ndc.csv, ARM_ndc.csv, ATG_ndc.csv, AUS_ndc.csv, AUT_ndc.csv; AZE_ndc.csv, BDI_ndc.csv, BEL_ndc.csv, BEN_ndc.csv, BFA_ndc.csv; BGD_ndc.csv, BGR_ndc.csv, BHR_ndc.csv, BHS_ndc.csv, BIH_ndc.csv. The files are represented by small icons and labeled with their respective country names.

The screenshot shows a Microsoft Excel spreadsheet titled 'ModelData'. The first row contains column headers corresponding to the variables: 'country', 'a_agriculture_a_coastal_rea_cross_cutt_a_dmn', 'a_education_a_energy', 'a_environment_a_health', 'A_im_Capita_A_im_Iran_A_im_TecTri_hufuf', 'a_social_deve_a_tourism', 'a_transport_a_urban', 'a_water_m_agriculture_m'. The data consists of 180 rows, one for each country, with numerical values for each variable. The cells contain either numerical values or 'NA' for missing data. The Excel ribbon at the top shows various tabs like Insert, Draw, Page Layout, Formulas, Data, Review, View, and Tell me.

Data Clean

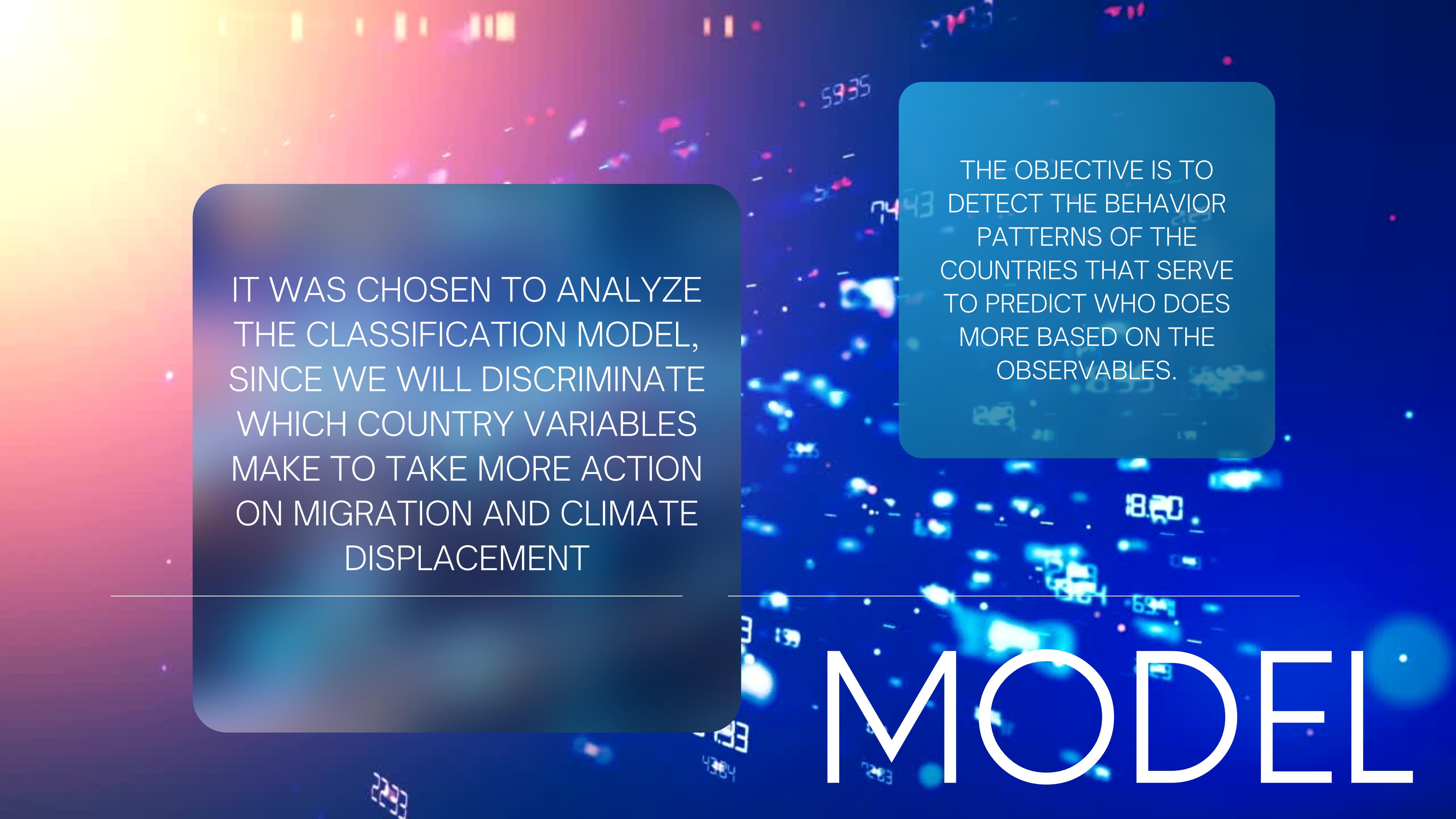
We follow the same procedure for sociodemographic information from the World Bank. We clean csv to decrease N/A and we put together indicators, NDC and the information of the World Bank in a new database.



New Base

It should be noted that the data obtained had to be transformed to qualitative, we have binary variables. Data were collected from 197 countries. This new base in our code was cleaned up again, obtaining data from 180 countries (numerical variables).





IT WAS CHOSEN TO ANALYZE THE CLASSIFICATION MODEL, SINCE WE WILL DISCRIMINATE WHICH COUNTRY VARIABLES MAKE TO TAKE MORE ACTION ON MIGRATION AND CLIMATE DISPLACEMENT

THE OBJECTIVE IS TO DETECT THE BEHAVIOR PATTERNS OF THE COUNTRIES THAT SERVE TO PREDICT WHO DOES MORE BASED ON THE OBSERVABLES.

MODEL

How is the model?

Migration and climate
displacement

Classification model



Countries
173 High and 7 Low

1.5 as threshold

**Interest response is
divided into two
groups**

We fit a classification tree to our data set response of interest "migration_and_displacement" in countries and classified the problem based on the mean of the interest response.

Where tresshold based on .10 of the observations that are below that limit, the threshold is equivalent to 1.5, we generated a binary response, saving it as a factor, if the response of interest is above this function has high mitigation actions to the response of interest.

The model we want to estimate
`model<-as.formula(paste0("response.binary","~",paste(pr
edictors,collapse="+)))`

- 
1. Regression Three
 2. Random Forest
 3. Linear Regression

VALIDATION

Classification tree

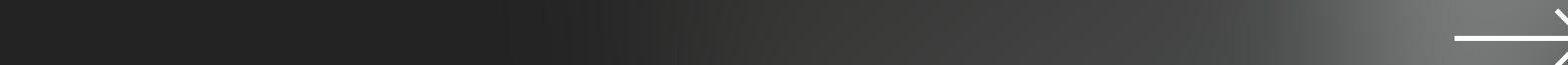
About it:

Train with sample of 100

5 nodes

Residual mean deviance: 0.1235

minor error: 59.86056



Total Population < 7.73684e+07

Barriers < 1.5

Water < 1

Low

High

Urban < 1

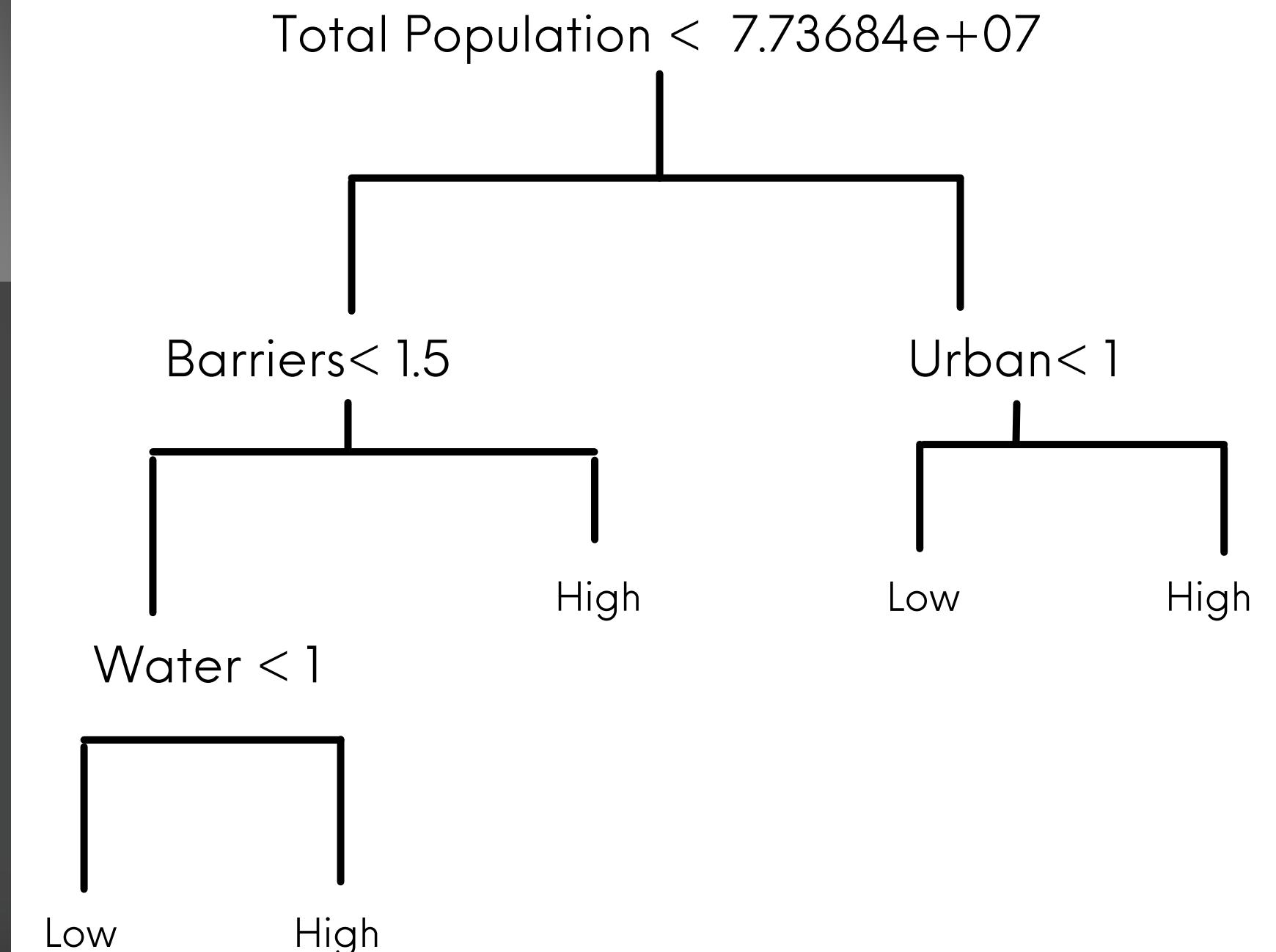
Low

High

What does it tell us?

Countries with large populations and where Urban has a score lower than one means that they have a comprehensive NDC in the urban part where the variable of migration and displacement is taken into consideration.

Countries with small populations and those with the most technical barriers in terms of climate policy implementation are the ones that are most concerned about the effect of migration and displacement. The countries with the fewest barriers regarding the implementation of climate policy are concerned about the effects of migration on the part of water resources.



Linear regression

About it:

Response: Migration_and_displacement.

Significant variables thrown by the classification tree.

Based on the p-value, the most important variables are the actions on water and urban for mitigate the effects of climate change in the phenomenon of migration and displacement.

Call:

```
lm(formula = migration_and_displacement ~ SP.POP.TOTL + M_PL7 +  
    a_water + a_urban, data = data.model)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.07387	-0.07962	-0.05944	-0.05072	1.94807

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.060e+00	7.057e-02	29.191	<2e-16 ***
SP.POP.TOTL	-1.508e-10	2.620e-10	-0.576	0.565
M_PL7	-1.110e-04	2.128e-02	-0.005	0.996
a_water	1.014e-02	8.194e-03	1.238	0.217
a_urban	-2.679e-02	2.674e-02	-1.002	0.318

Signif. codes:

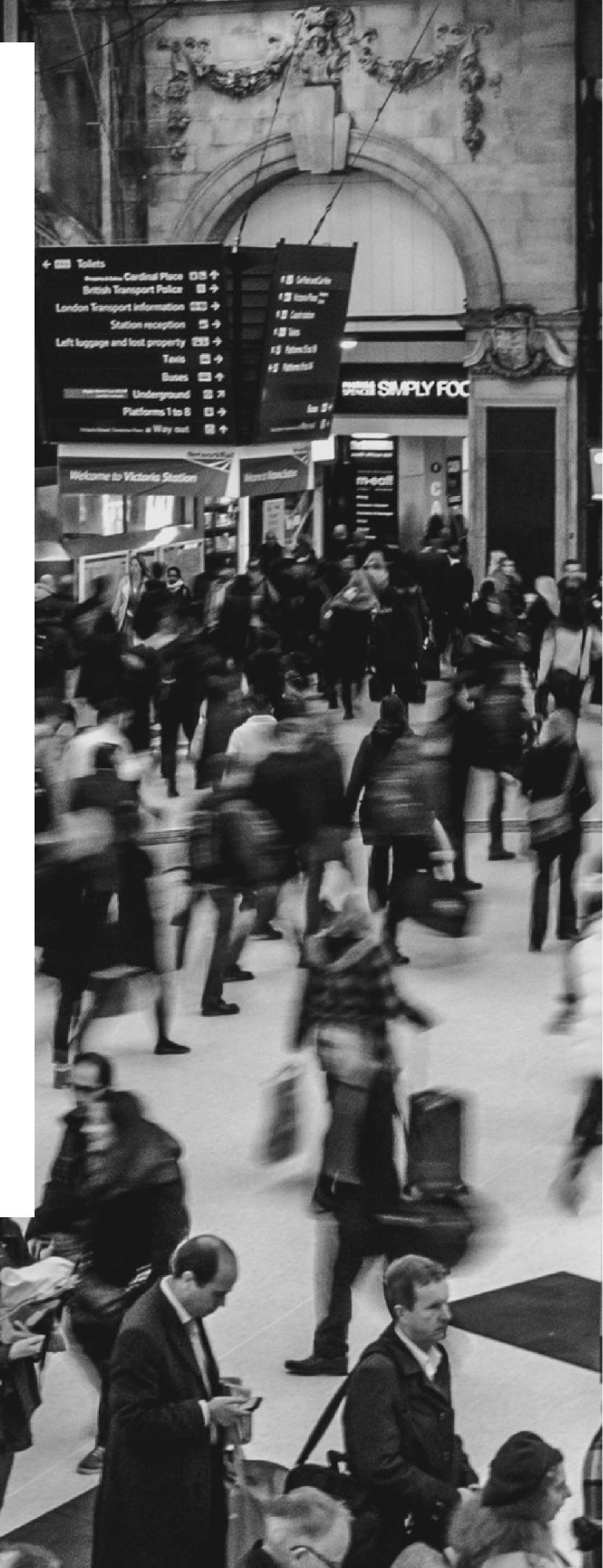
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

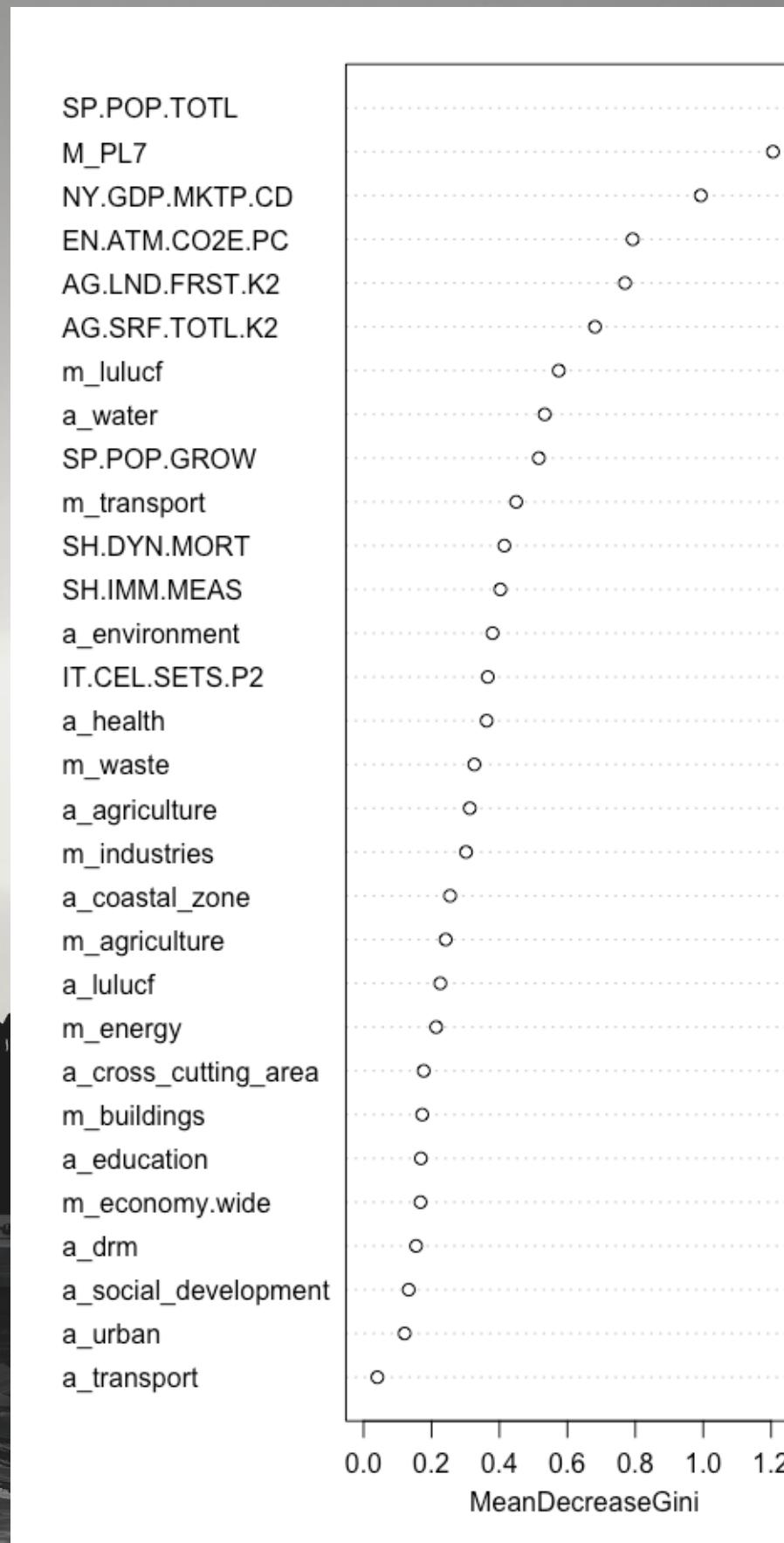
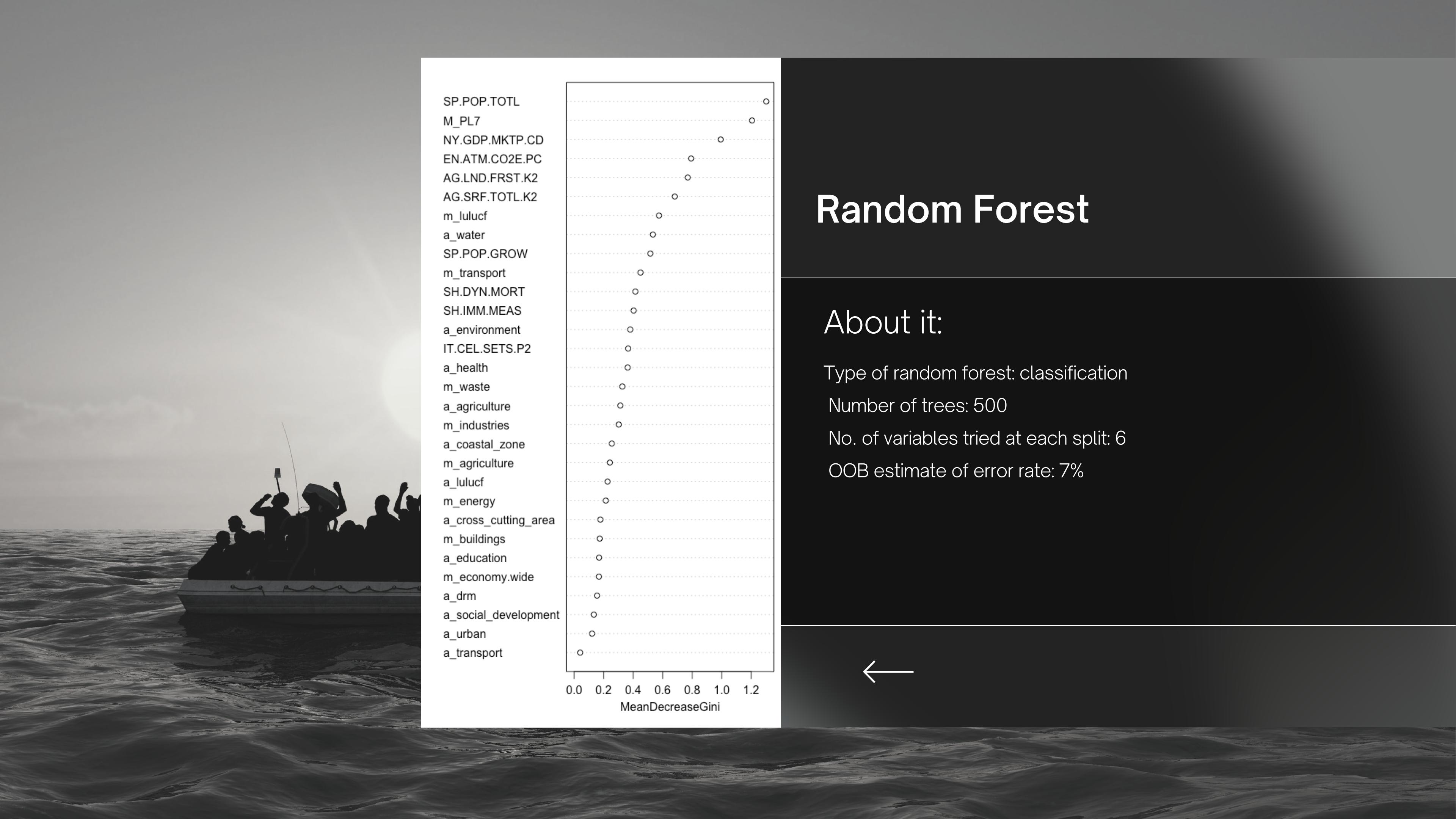
Residual standard error: 0.5084 on 184 degrees of freedom

(5 observations deleted due to missingness)

Multiple R-squared: 0.01144, Adjusted R-squared: -0.01005

F-statistic: 0.5323 on 4 and 184 DF, p-value: 0.7122





Random Forest

About it:

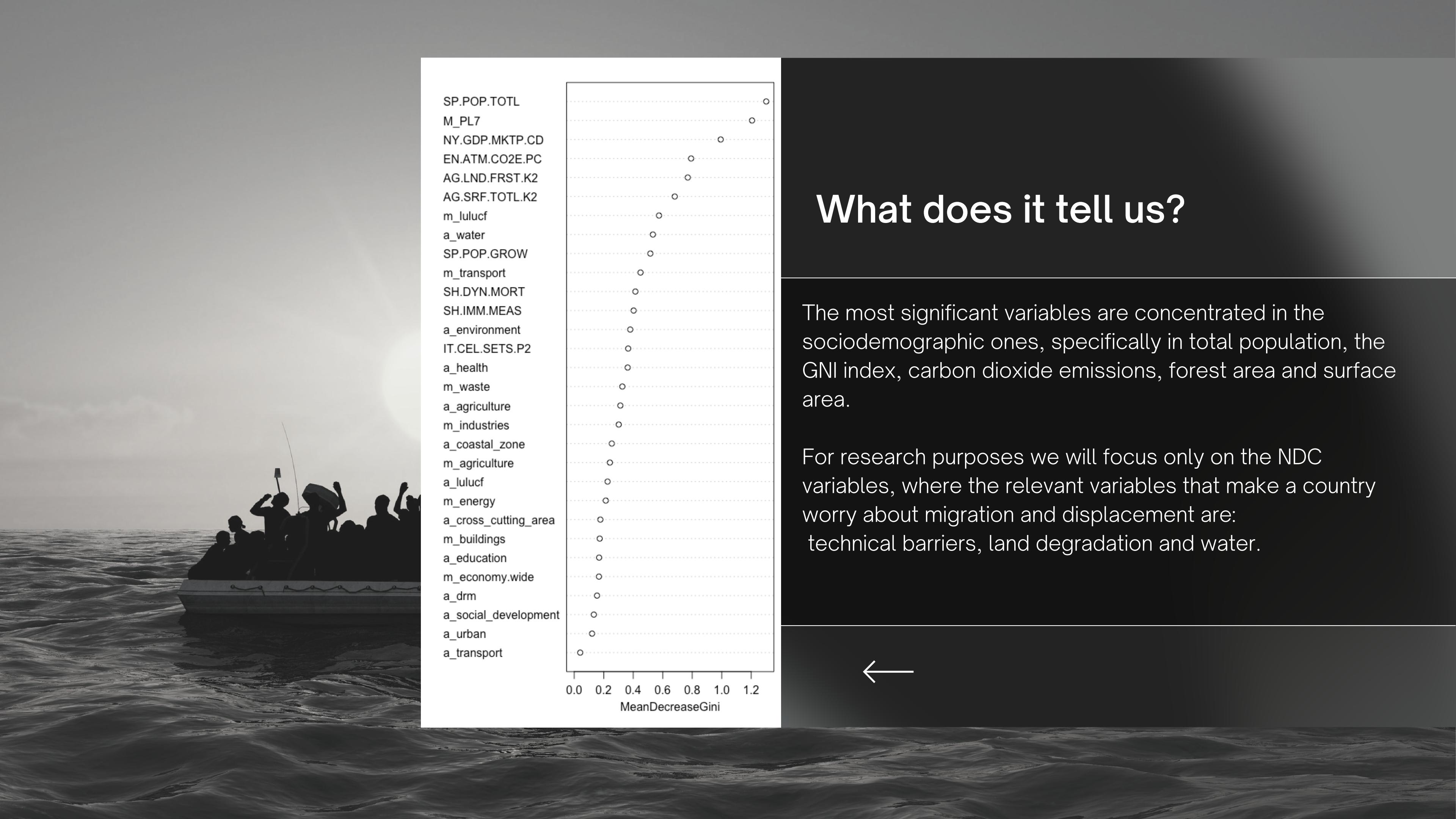
Type of random forest: classification

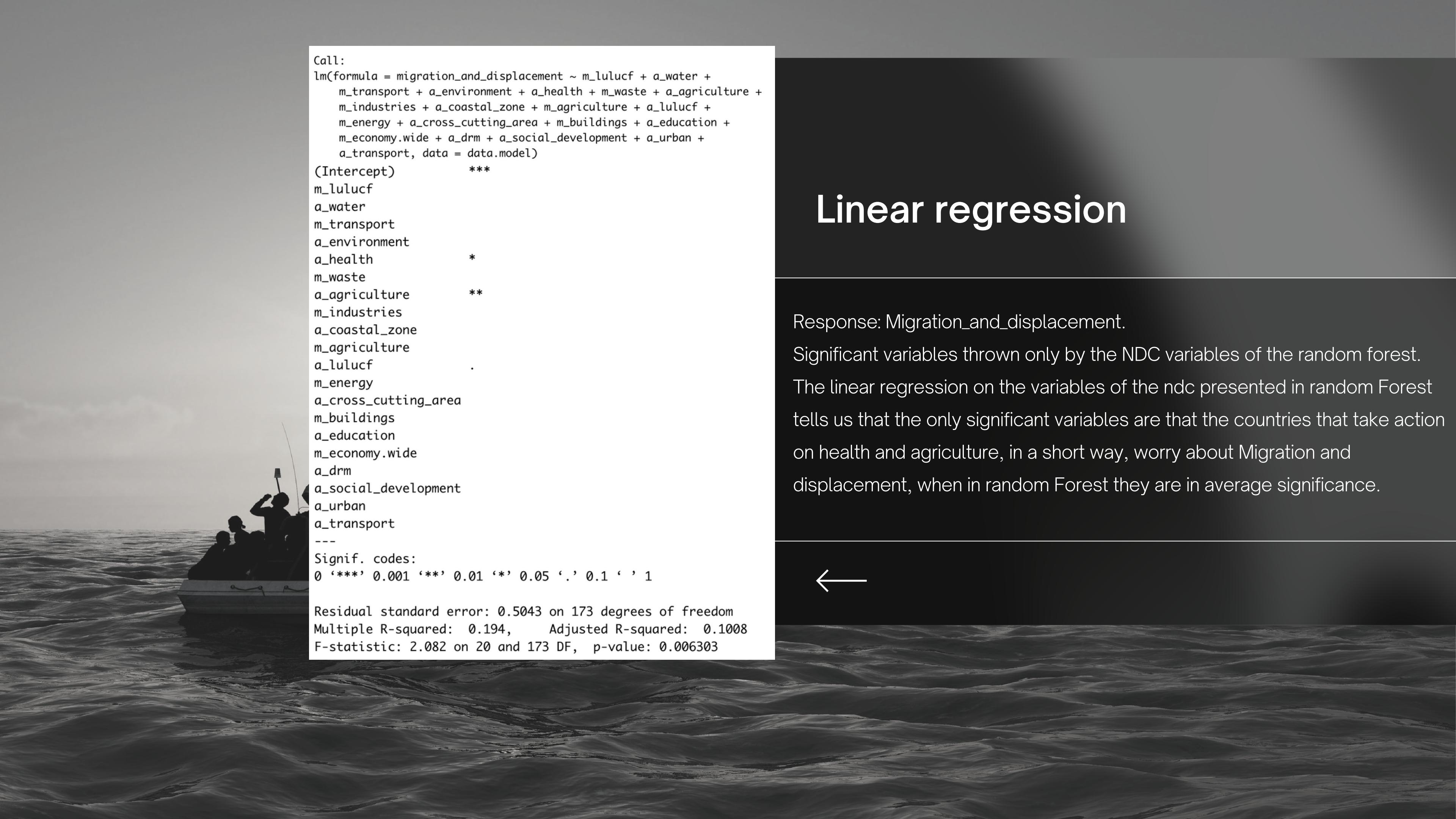
Number of trees: 500

No. of variables tried at each split: 6

OOB estimate of error rate: 7%







```
Call:  
lm(formula = migration_and_displacement ~ m_lulucf + a_water +  
    m_transport + a_environment + a_health + m_waste + a_agriculture +  
    m_industries + a_coastal_zone + m_agriculture + a_lulucf +  
    m_energy + a_cross_cutting_area + m_buildings + a_education +  
    m_economy.wide + a_drm + a_social_development + a_urban +  
    a_transport, data = data.model)  
(Intercept) ***  
m_lulucf  
a_water  
m_transport  
a_environment  
a_health *  
m_waste  
a_agriculture **  
m_industries  
a_coastal_zone  
m_agriculture  
a_lulucf .  
m_energy  
a_cross_cutting_area  
m_buildings  
a_education  
m_economy.wide  
a_drm  
a_social_development  
a_urban  
a_transport  
---  
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 0.5043 on 173 degrees of freedom  
Multiple R-squared: 0.194, Adjusted R-squared: 0.1008  
F-statistic: 2.082 on 20 and 173 DF, p-value: 0.006303
```

Linear regression

Response: Migration_and_displacement.

Significant variables thrown only by the NDC variables of the random forest.

The linear regression on the variables of the ndc presented in random Forest tells us that the only significant variables are that the countries that take action on health and agriculture, in a short way, worry about Migration and displacement, when in random Forest they are in average significance.



Findings | Significant variable for Migration and Displacement

Why do some commit to doing more and others less?

Findings



Water



Barriers



Health



Agriculture



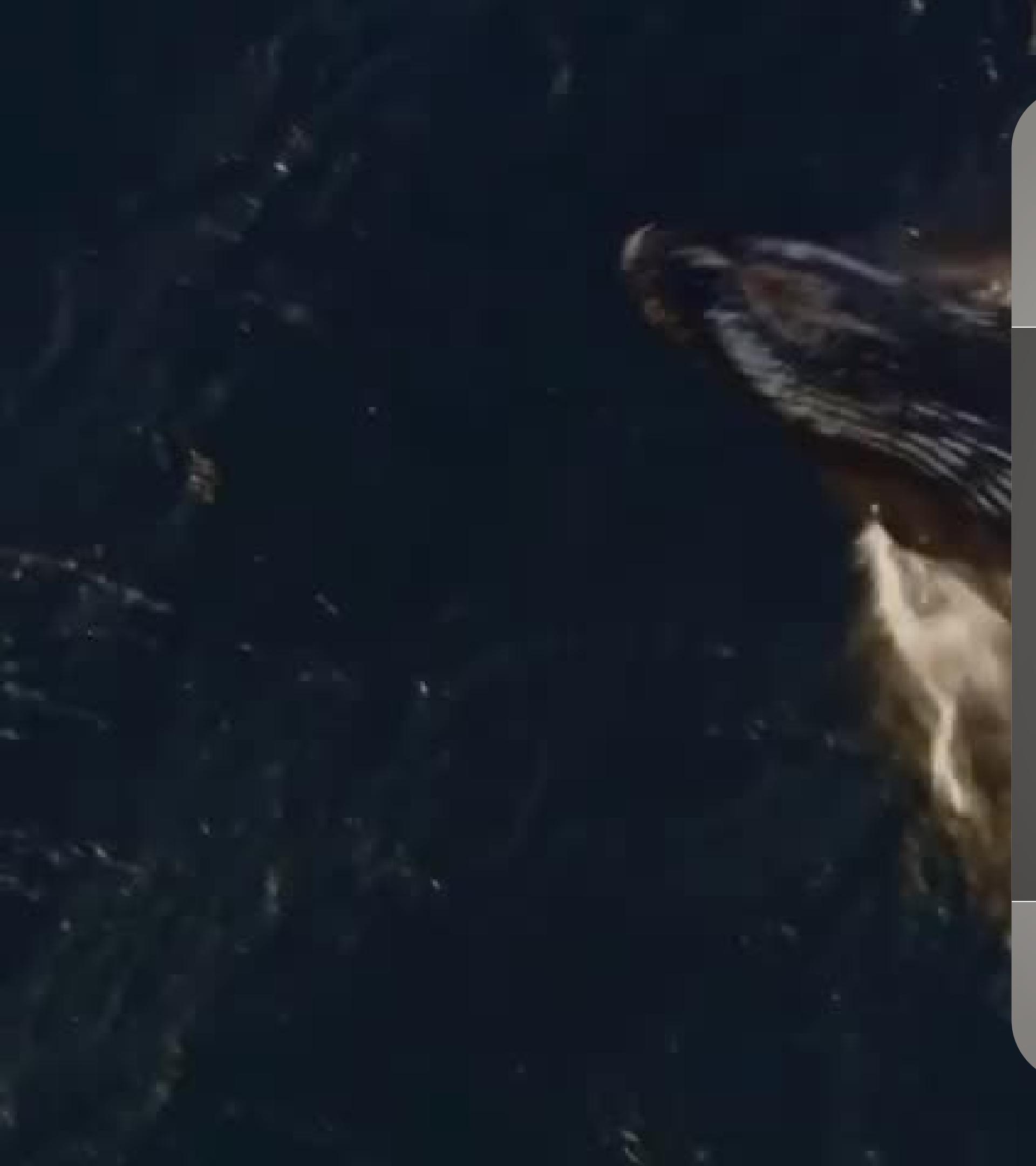
Total Population

The coincidences of Random Forest, Classification tree and linear regressions were taken.
Land degradation was chosen to considering on agriculture for this purpose of analysis.





HOW TO
ACCELERATE
CLIMATE ACTION?



Biomimetic

Biomimetic, is the practice of developing sustainable human technologies inspired by nature.

It's a science which needs to go hand in hand with political decision makers to implement solutions that will lead to more environmentally friendly nature-based adaptation strategies that focus on regeneration and restoration



Plan and program

Integrate climate migration and displacement in planning and development, to promote innovative solutions through technology.

to promote environmental innovation we have to have a new environment of public policies focused on biomimetics

Public Policy

Catalyze climate innovation by investing in new technologies

Invest and potentiate biomimetics to take care of the significant variables: water, agriculture and health, through existing innovative projects and promote constant innovation of new projects to bring them to reality, this science lacks investment and attention because it is emerging and little known.

Investment



A large, detailed Earth globe is positioned in the center of the frame, resting on a calm, blue ocean. The globe shows the continents of North America and South America. The background is a bright blue sky with scattered white and light blue clouds. A thin horizontal black bar is overlaid across the middle of the image, containing the text.

"An environmental problem needs sustainable solutions"

Thanks for you attention!