

Population Association of America 2021 Annual Meeting



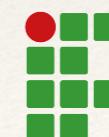
Data Visualization for Epidemiological and Demographic Data for Malaria Surveillance in the Brazilian Amazon, 2007-2019

Natália Arruda
natiarruda@gmail.com

Vinícius de Souza Maia
viniciusmaia108@gmail.com

Luciana Alves
lcalves@unicamp.br

LADDEM (Data Analysis in Demography Lab) - NEPO/UNICAMP
Population Studies Center Elza Berquó
Institute of Philosophy and Human Sciences
University of Campinas - Brazil



INSTITUTO FEDERAL
São Paulo
Câmpus Campinas



Background

- ❖ B.A. Informatics (2002), MSc. in Mechanical Engineering (2006), more than 15 years experience with Databases Development.
- ❖ Computer Science Professor @ *Federal Institute of São Paulo - IFSP*
- ❖ Ph.D Candidate @ *Demography Post Graduation Program - University of Campinas - UNICAMP Supervisor:* PhD Luciana Correia Alves
- ❖ Data Science Research
 - ❖ *2018 - 2020: Decision-Making Support Platform Based on Visual Analytics and Machine Learning to Subsidize Public Politics Focused on Gestational Health (Bill & Melinda Gates Foundation grant)*
 - ❖ *2020 - 2022: Data Science applied to epidemiological and demographic information as a strategy to simulation and malaria vigilance monitoring in the Brazilian Amazon (Bill & Melinda Gates Foundation grant)*

Acknowledgment

DISQUE
SAÚDE
136



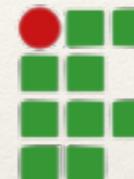
MÍNISTÉRIO DA
SAÚDE



Conselho Nacional de Desenvolvimento
Científico e Tecnológico



BILL & MELINDA
GATES foundation



INSTITUTO FEDERAL
São Paulo
Câmpus Campinas



home visualizations predictions discoveries upload your data

We provide demographic, socioeconomic, maternal and child health indicators at the federal, state and municipal levels.

Our main proposal is to provide managers and society as a whole with useful data for the formulation of prevention policies and actions, as well as contributing to the improvement of the quality of health care and the information generated at the basic levels of care, with consequent reduction of neonatal mortality.

GO TO VISUALIZATIONS

This work is supported by:

- Bill & Melinda Gates Foundation [OPP1201970];
- Brazilian Ministry of Health, Brazilian National Council for Scientific and Technological Development [443774/2018-8];
- NVIDIA corporation for supporting our research with a TITAN Xp GPU donation;
- Center for Information Technology Renato Archer (CTI).

<http://www.plataformasami.com.br>

Introduction

- ❖ Malaria is one of the main public health problems worldwide
- ❖ According to World Malaria Report:
 - ❖ 2017 - 231 mi cases and 416k deaths
 - ❖ 2018 - 228 mi cases and 405k deaths
- ❖ Mortality caused by Malaria has direct impact for some Demography studies

Introduction

- ❖ Still a huge challenge in Brazil
 - ❖ 34.4% of cases registered in the American continent
 - ❖ After 1940 it was almost eliminated from extra-Amazonian region
 - ❖ Currently, 99% occurs in Legal Amazonia
 - ❖ 1998–2000: 600k cases
 - ❖ 2017: 194k cases

Objectives

- ❖ Provide resources for monitoring, measurement and evaluation of the epidemiological / demographic profile of Malaria
 - ❖ Develop tools to support health managers, with information enabling a better understanding of the problem
 - ❖ Explore SIVEP-Malaria dataset (Brazilian Amazon 2007-2019)
 - ❖ Design of interactive data visualizations related to Malaria Surveillance
- ❖ This project is aligned with the **Sustainable Development Goals** and with **Global technical strategy for Malaria 2016–2030**

SIVEP-Malaria Dataset

- ❖ Platform for collecting and disseminating Malaria surveillance data in Brazil

Table 1. SIVEP data description table - Adapted from WIEFELS et al. (2016)

Variável	Definição	Variável	Definição	Variável	Definição	Variável	Definição
COD_NOTI	Notification number	DT_NASCI	Birth date	MUN_RESI	Municipality of residence	LOC_INFE	Locality of infection
DT_NOTIF	Notification date	ID_PACIE	Patient age	LOC_RESI	Locality of residence	DT_EXAME	Examination date
TIPO_LAM	Active/passive	ID_DIMEA	Age writing format	DT_SINTO	First symptoms date	EXAME	Examination method
UF_NOTIF	State of notification	SEXO	Sex	DT_TRATA	Date of treatment	RES_EXAM	Examination results
MUN_NOTI	Municipality of notification	GESTANTE	Pregnancy length	VIVAX	Patient is under Vivax treatment	QTD_CRUZ	Parasitaemia
COD_UNIN	Health unit of notification	NIV_ESCO	Schooling level	FALCIPARUM	Falciparum treatment	QTD_PARA	Parasites by mm ³
COD_AGEN	Health agent code	RACA	race/skin color of the patient	ID_LVC	Follow-up consultation	HEMOPARASI	Hemoparasites
SEM_NOTI	Notification week	COD_OCUP	Employment	PAIS_INF	Country of infection	EXAMINADOR	Examiner code
DT_DIGIT	Date of digitalization	PAIS_RES	Country of residence	UF_INFEC	State of infection	TREATMENT_SCHEDULE	Treatment schedule
DT_ENVLO	Data entering into National database date	UF_RESID	State of residence	MUN_INFE	Municipality of infection	SINTOMAS	Symptoms
Administrative data		Patient data		Epidemiological and laboratorial data			

SIVEP-Malaria Dataset

- ❖ The dataset built for this study has information from 2007 to 2019

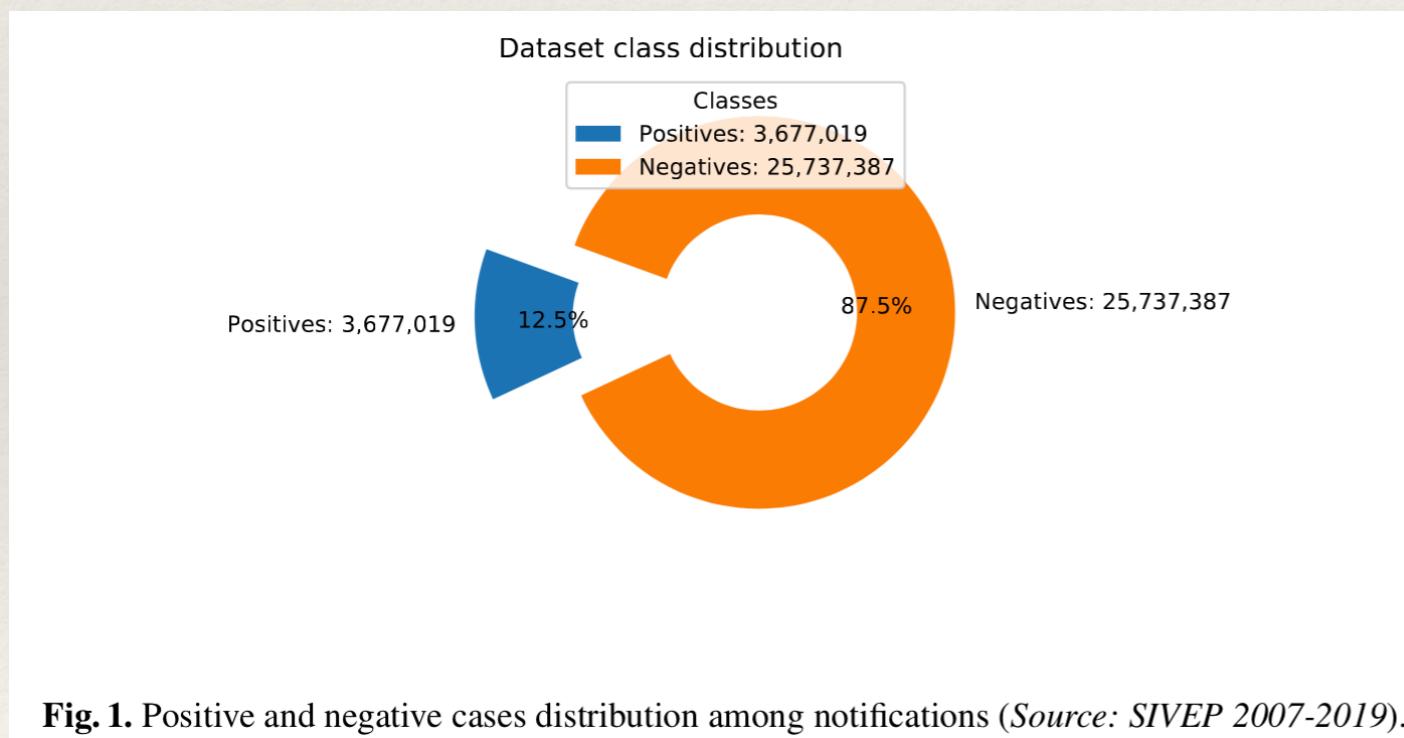


Fig. 1. Positive and negative cases distribution among notifications (Source: SIVEP 2007-2019).

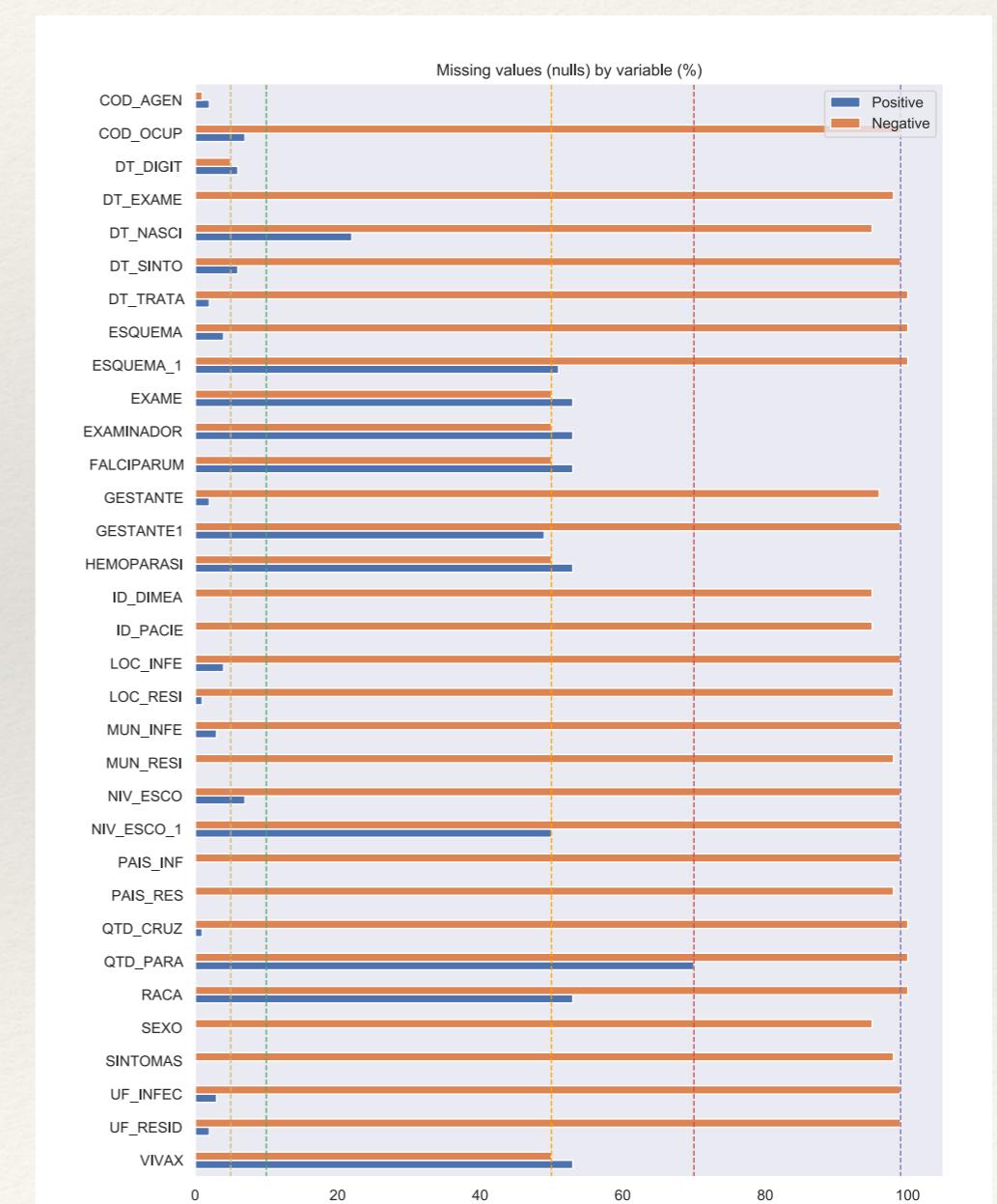


Fig. 2. Percentage of null values on each variable on the dataset sample used on this study (Source: SIVEP 2007-2019).

Dataset Preprocessing

- ❖ Missing or inconsistent data in Brazilian public health repositories is very common
- ❖ Mostly due incorrect filling of handwritten forms
- ❖ To tackle this problem the following were performed:
 - ❖ Excluding records reporting age older than 105 years;
 - ❖ Excluding records indicating pregnancies on women younger than 10 years of age;
 - ❖ Excluding records with incoherent variable combinations

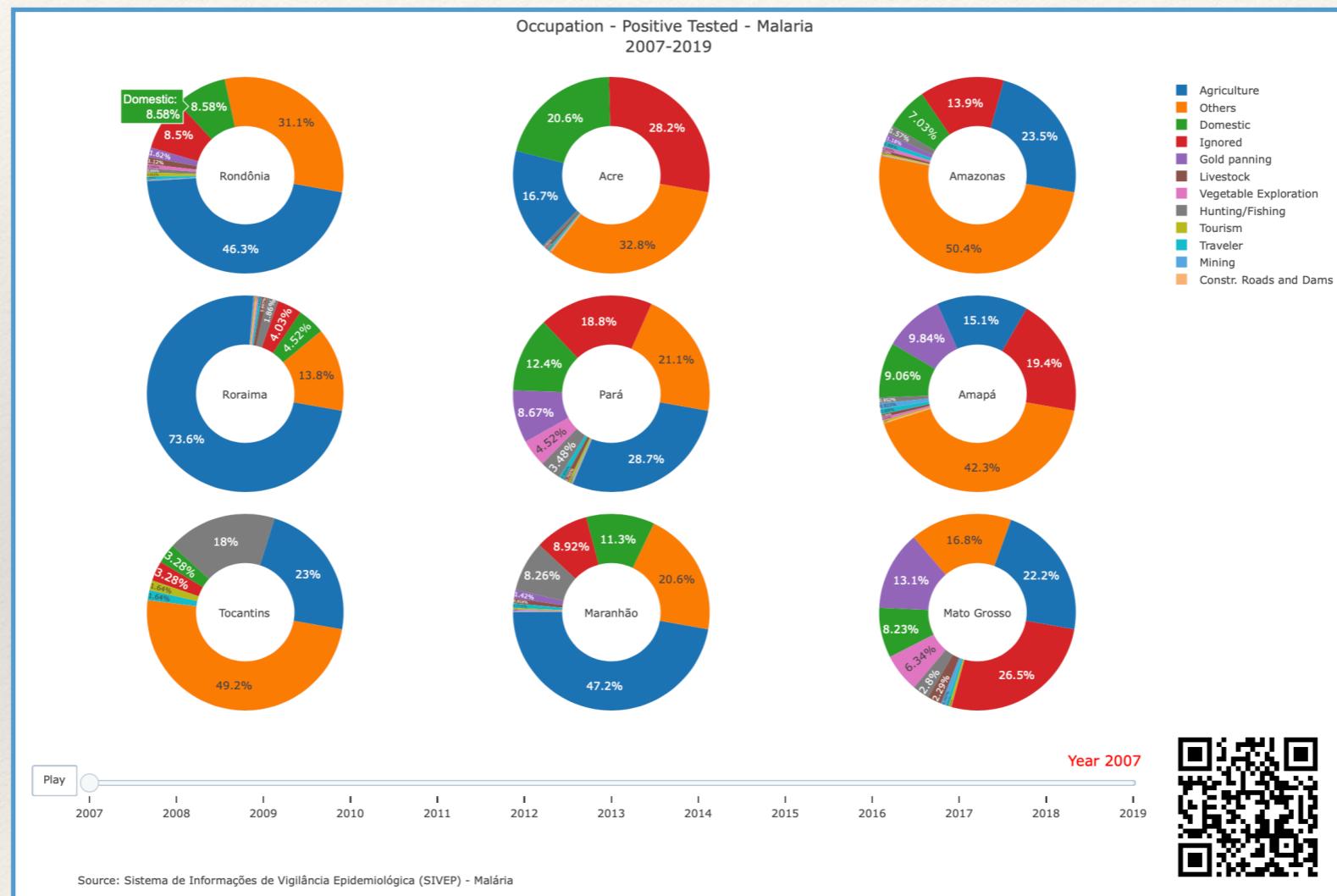
Methods

- ❖ Knowledge discovery approach
 - ❖ Obtaining relevant information group by cases (positive or negative)
 - ❖ Allow exploration of specific aspects of each class
 - ❖ Developed using R, Plotly, ggplot2 and Shiny
- ❖ Data was extracted from the pre-processed dataset in CSV
- ❖ Grouping and summarizations performed using R scripts

Dataviz 3.1 - Demographic/Socioeconomic Profile

https://arrudanat.shinyapps.io/ocup_eng

- ❖ Consolidates the occupational profile of infected individuals

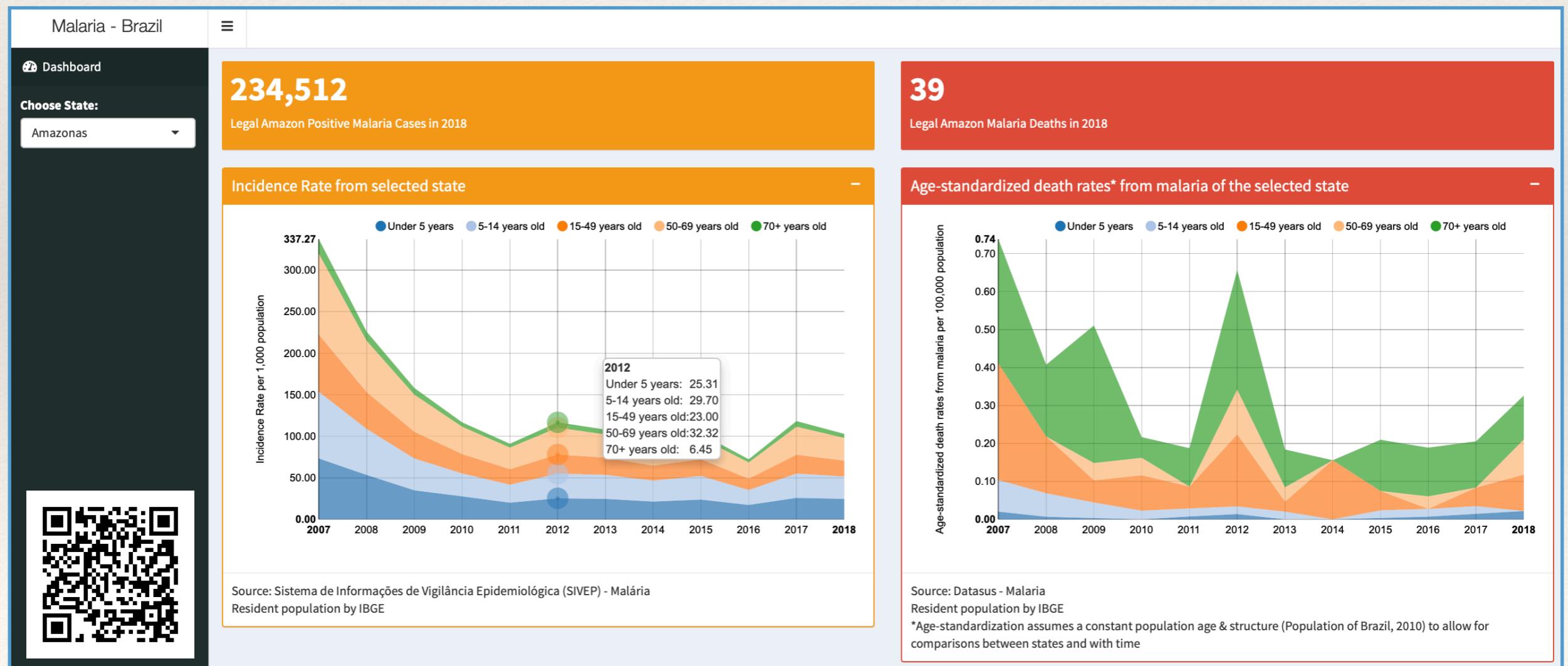


Select a year and compare among all Legal Amazon Federative Units, the distribution of the main activity performed by the patient in the last 15 days according to areas at risk of catching Malaria.

Dataviz 3.2 - Incidence and death rates

<https://arrudanat.shinyapps.io/taxasMalaria>

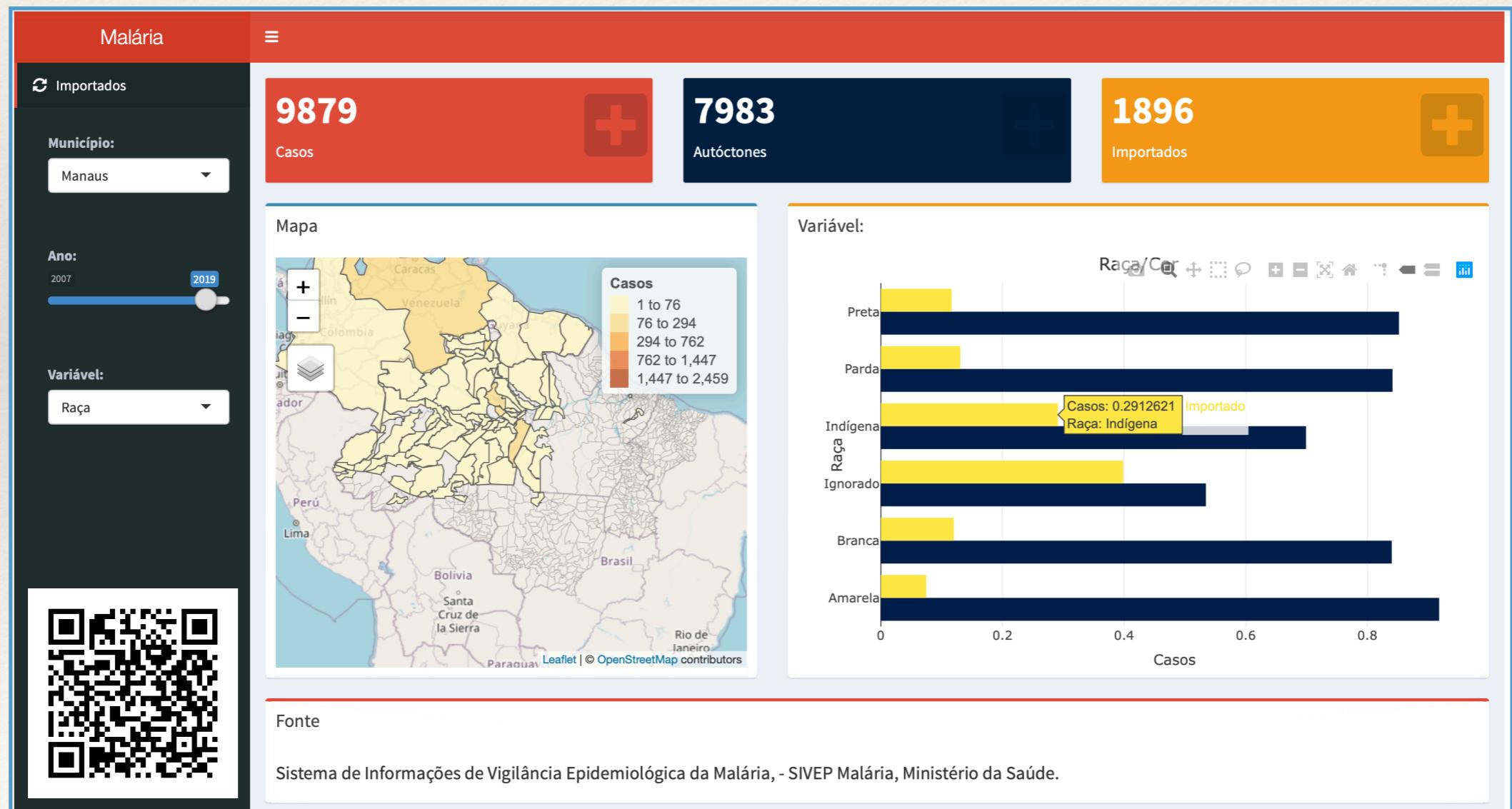
- Shows incidence and mortality rates, by age group and year of notification, and the user can select the Federative Unit



Dataviz 3.3 - Imported vs Autochthonous

<https://vinicius-maia.shinyapps.io/importadosdash>

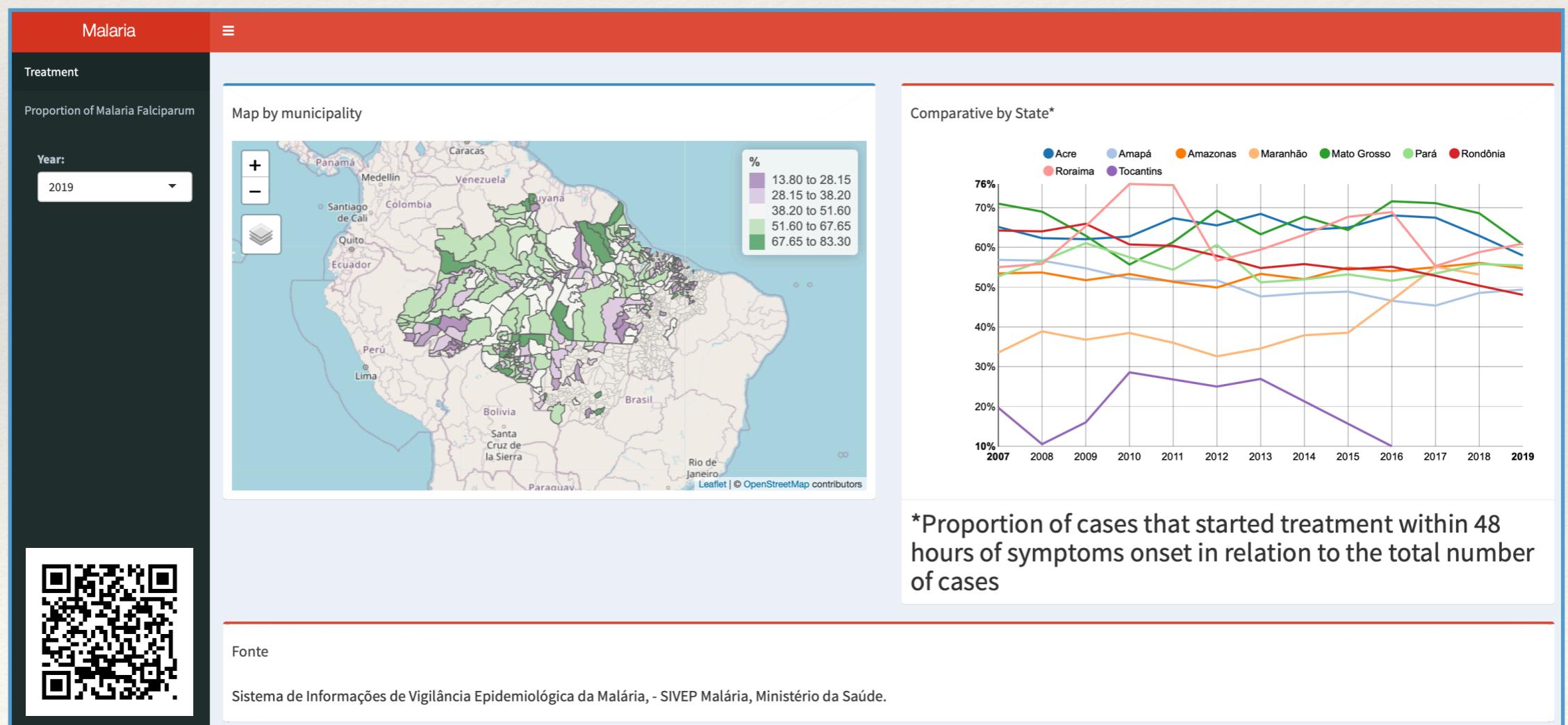
- ❖ Distribution between imported and autochthonous Malaria cases
- ❖ User can select municipality, year and the variable of interest



Dataviz 3.4 - Treatment within 48 hs

https://arrudanat.shinyapps.io/IndicadoresProcesso_eng

- ❖ 3 different groups of information using heat map combined with line graph
 - ❖ Proportion of positive cases that started treatment within 48 hs
 - ❖ Percentage of infections by Plasmodium Falciparum
 - ❖ Percentage of infections by Plasmodium Vivax



Discussion

- ❖ WHO continues to highlight the urgent need for new and improved tools in global response to Malaria
 - ❖ Surveillance is therefore the basis of operational activities in settings of any level of transmission
 - ❖ Its objective is to support reduction of the burden of the disease, eliminate it and prevent its re-establishment.
 - ❖ We hope tools of this kind can reduce the barrier to data use in policy decisions as it reduces the burden of data extraction and analysis on health staff and local policy makers.

Machine learning for malaria treatment schema recommendation using routine surveillance data

Carlos Eduardo Beluzo, Everton Josué da Silva, Natália Martins Arruda, Vinícius de Souza Maia, Bianca Cechetto Carlos, Tiago Carvalho and Luciana Correia Alves

Abstract

Malaria, a parasitic disease predominating in less developed regions, is a worldwide public health problem, affecting some of the largest populations in the world. Among the various factors that favor response actions to infected individuals, the treatment schema selected to be used in infected individuals are crucial for the effective and efficient recovery of patients, as well as for public health planning activities. According to information from Brazilian Malaria Epidemiological Surveillance System (SIVEP-Malaria), currently, there are approximately 22 treatment schemes that can be recommended to patients, and their selection is determined by several factors. In this paper we present a machine learning model to recommend which treatment to use in patients diagnosed with malaria. The model is trained with epidemiological, demographic and socioeconomic information. In preliminary experiments, using the Extreme Gradient Boost algorithm, two models were built, for the two most frequent treatment schemas. The models achieved an accuracy value of 0.94 with an AUC value of 0.914 when predicting “Schema treatment 6” against all other schemas, and accuracy value of 0.93 with an AUC value of 0.911 when predicting “Schema treatment 17”.



Population Association of America 2021 Annual Meeting

Machine Learning for neonatal mortality rate forecasting using public health data from São Paulo city (2006-2016)

Carlos Ed Beluzo
linkedin.com/in/cbeluzo
cbeluzo@gmail.com

Natália Arruda | Luciana Alves | Cátia Sepetauskas | Tiago Carvalho



Methods

❖ Database

- ❖ Model relays on knowledge learned from a database with more than 28mi individuals (Brazil 2006-2016¹)
- ❖ Health access variables gathered from the National Register of Health Facilities
- ❖ Health Region Code used as key to build a monthly bases time series of 10 years for each health regions in Brazil (438 regions)
- ❖ Features selected based on feasibility to be managed by policy makers in real world

<i>Past NMR (per month)</i>	<i>Number of mothers having schooling between 8 to 11 years</i>	<i>Number of health care units</i>	<i>Number of neonatal incubators</i>	<i>Number of newborn intensive care units</i>
---------------------------------	---	--	--	---

¹ <http://dx.doi.org/10.18063/ijps.v5i2.1130>

Results

- ❖ Low accuracy on absolute values for NMR
- ❖ Prediction curves are capable of indicating NMR trends
- ❖ New results was posted at **MedRxiv** on April 25, 2021¹

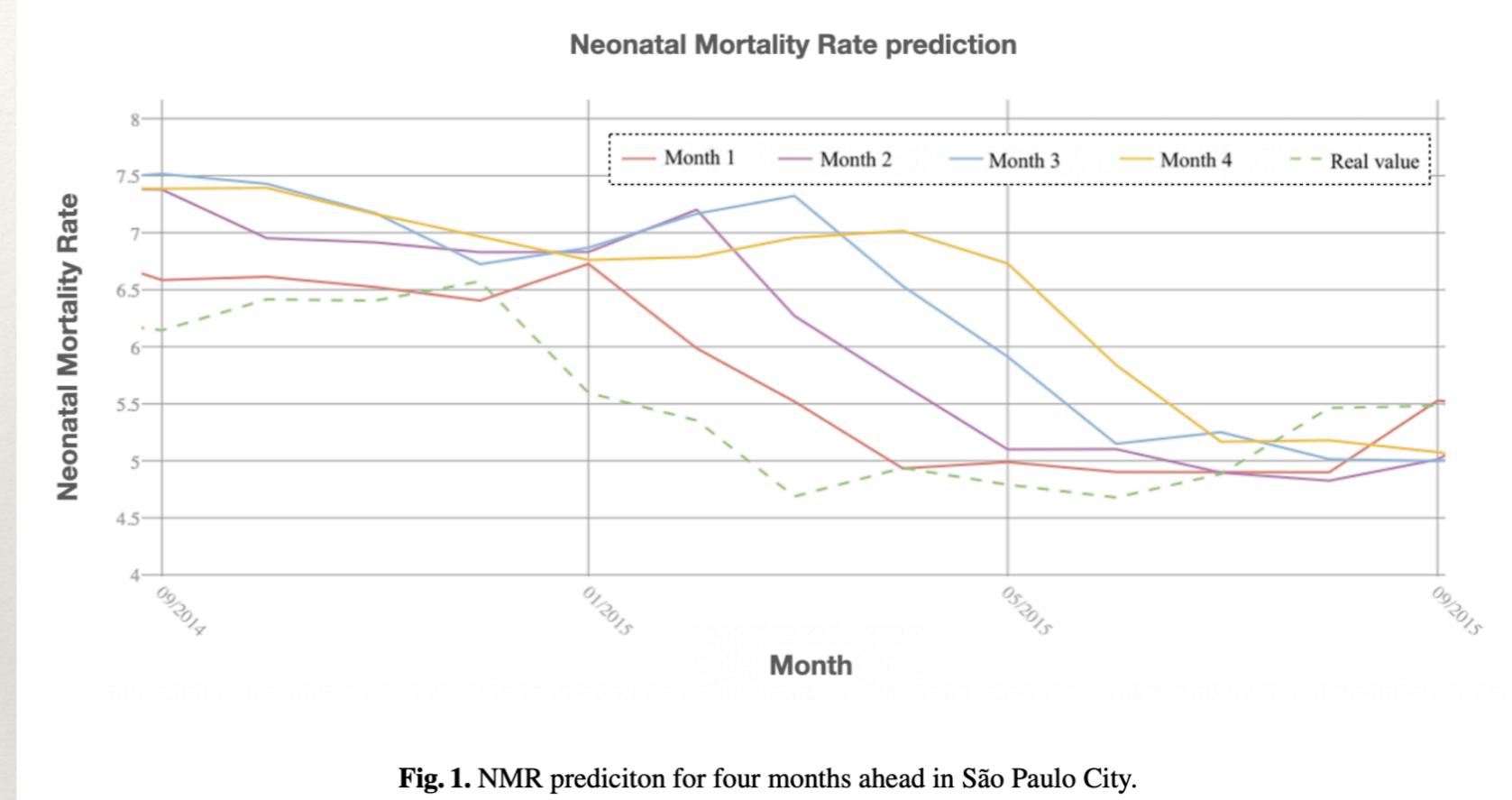


Fig. 1. NMR predictiton for four months ahead in São Paulo City.

Months ahead	1			2			3			4		
Metric	MSE	RMSE	MAPE	MSE	RMSE	MAPE	MSE	RMSE	MAPE	MSE	RMSE	MAPE
Value	0.169	0.411	5.338	0.428	0.654	8.664	0.63	0.789	10.768	0.71	0.842	11.632

Table 1. Error metrics for projections of 1, 2, 4 and 4 months ahead of the NMR.

¹ <https://doi.org/10.1101/2021.04.22.21255916>



NeMoR: a New Method Based on Data-Driven for Neonatal Mortality Rate Forecasting

 Comment on this paper

 Previous

Posted April 25, 2021.

 Carlos Eduardo Beluzo,  Luciana Correia Alves,  Natália Martins Arruda,  Cátia Sepetauskas,
 Everton Silva,  Tiago Carvalho

doi: <https://doi.org/10.1101/2021.04.22.21255916>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Abstract

Full Text

Info/History

Metrics

 Preview PDF

 Download PDF

 Email

 Author Declarations

 Share

 Data/Code

 Citation Tools

 XML

 Tweet

 Curtir 0

Thanks

cbeluzo@gmail.com