Big Data Visualization Methods Applied in the Context of Neonatal Mortality

Analysis of the Problem and its Causes

Carlos Eduardo Beluzo
Instituto Federal de São Paulo – IFSP
beluzo@ifsp.edu.br

Lucas Rodrigues Pimentel
Instituto Federal de São Paulo – IFSP
l.pimentel@aluno.ifsp.edu.br

Tiago José de Carvalho
Instituto Federal de São Paulo – IFSP
tiagojc@ifsp.edu.br

ABSTRACT

Neonatal mortality is a worldwide problem and its reduction is of common international interest. Around 46% of all deaths in the world happen to infants up to five years of age, and most of those deaths is concentrated in the first few days after birth. In Brazil, although having data sources such as IBGE and DATASUS, there is no platform that provides an intuitive way of visualizing this data. Thus, this work proposes the use of specialized visual analysis tools for big data, such as Python libraries Dash by Plotly, NumPy and Pandas, in order to facilitate the understanding of public information related to neonatal mortality, which can lead to the mitigation of the problem.

KEYWORDS

Neonatal Mortality, Data Sources, Big Data, Visualization

1 INTRODUCTION

Reducing neonatal mortality is of paramount importance to the whole world. The neonatal period is the most vulnerable moment in the life of a child and it is heavily influenced by living conditions of the population and health care. Brazil, for example, has worrying levels of neonatal mortality. Although there are different public data sources such as IBGE and DATASUS, relying on those databases without being able to properly visualize the data is pointless. DATASUS is a good example of a big data source, and although being rich in relevant information, they are not used as much as they could be due to the difficulty of extracting the information and the lack of specialized tools to visualize them. This work proposes the design of interactive visualizations of data related to neonatal mortality using the Python 3 programming language.

2 MATERIALS AND METHODS

Gathering mortality and birth data is an important task and, in Brazil, the systems adopted to collect this type of data are the Mortality Information System (SIM) and the Born Alive Information System (SINASC). These systems have been greatly improved since their creation, reaching great coverage and data quality, making these sources reliable and accurate enough. Besides SIM and SINASC, there are other databases, such as DATASUS [1] and IBGE [2], that provide indicators related to neonatal mortality and applied to construct visualizations. In this work, the main indicators on which visualizations have been based are the Human Development Index (HDI), Gross Domestic Product (GDP), number of prenatal medical checkups and relation between types of birth (caesarean or vaginal delivery) and mortality. The methodology proposed for this work consists of implementing interactive visualizations of the data related to neonatal mortality, making the access to information simple and accessible to users. This has been done using the Python 3 programming language [3] and libraries such as Pandas [4], NumPy [5] and Dash by Plotly [6].

3 THE VISUALIZATIONS

Three different visualizations were made for this work using data provided by IBGE, DATASUS, SIM and SINASC. Data has been loaded using Pandas where major pre-processing has been applied, and then passed to Dash to plot the data. The indicators used can be split into two categories: per federative unit and health area[1]. The indicators with information related to federative unit are the proportion of caesarean births, proportion of hospital births, neonatal mortality, human development index and gross domestic product and diseases with highest lethality. The indicators with information related to health areas are the proportion of pregnant of women who started prenatal care in the first trimester of pregnancy.

3.1 Relation between birth types and mortality

The first visualization shows the relation between the proportion of birth types and neonatal mortality. For this visualization, the chosen plot was a simple scatter plot, and the types of birth were hospital and caesarean. The expected result was the decrease of mortality and increase in hospital births over time. The actual results, as seen in figures 1 and 2, depict a situation somewhat close to what was expected, although the increase in hospital births is subtle. Note that each circle represents a brazilian federative unit, and the neonatal mortality of it is represented by its radius. The horizontal axis is the proportion of caesarean births and the vertical axis is the proportion of hospital births.

---

[1]Continuous geographic space made of groups of neighboring municipalities, with the purpose of integrating the organization, planning and execution of health actions and services.
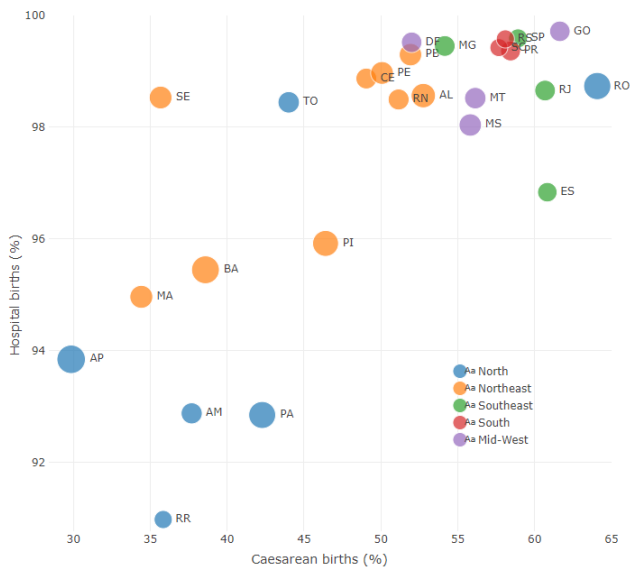
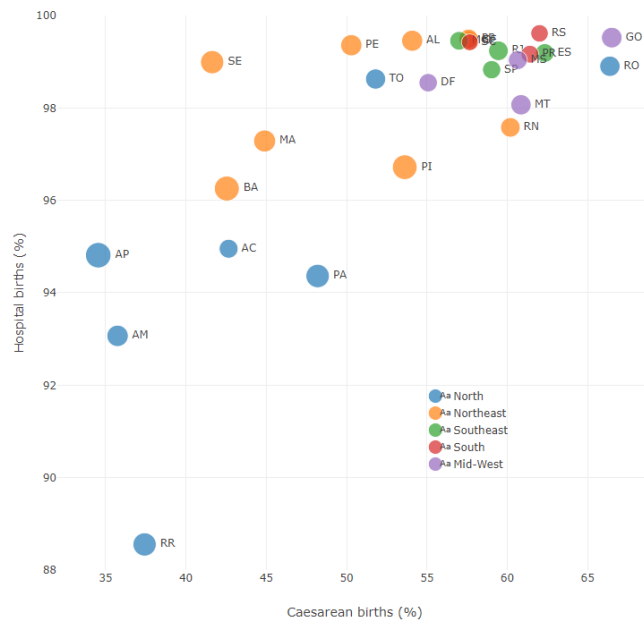**Figure 1: Relation between caesarean births, hospital births and neonatal mortality in 2010**

Figure 2: Relation between caesarean births, hospital births and neonatal mortality in 2016

It is possible to see that, as time passed and healthcare improved in Brazil, neonatal mortality decreased, caesarean births became more frequent and hospital births increased among many federative units, specially in the Northeast and North regions. However, in some specific locations, such as Roraima (RR), Amapá (AP) and Amazonas (AM), the number of hospital births slightly decreased,

but the neonatal mortality rate has decreased as well. The highest neonatal mortality recorded for 2010 was in Sergipe (SE), with 18.8 deaths per thousand births. The lowest was in Rio Grande do Sul (RS), with 9.5 deaths per thousand births. In 2016, the highest neonatal mortality rate recorded was 12.2 per thousand births in Amapá (AP), and the lowest was 6.3 in Santa Catarina (SC).

3.2 Relation between mortality, GDP and HDI

The second visualization proposed in this paper presents the relation between neonatal mortality (including a subdivision by its top 5 mortality causes), gross domestic product and human development index for any chosen brazilian federative unit. A donut chart was chosen for this visualization. The expected outcome was a significant decrease of neonatal mortality as GDP and HDI increased. Unfortunately, this hypothesis only remains valid in some federative units. In figures 3 and 4, despite total mortality rate remaining stable, it is easy to realize that newborn respiratory distress syndrome, congenital malformation of heart and extreme immaturity of newborn, which are 3 of the 5 main death causes, decreased along the years while the other two causes remained relatively stable. The same behaviour happened for bacterial sepsis of newborn and newborn respiratory distress, which decreased in São Paulo (SP), while the other causes kept reasonably stable.
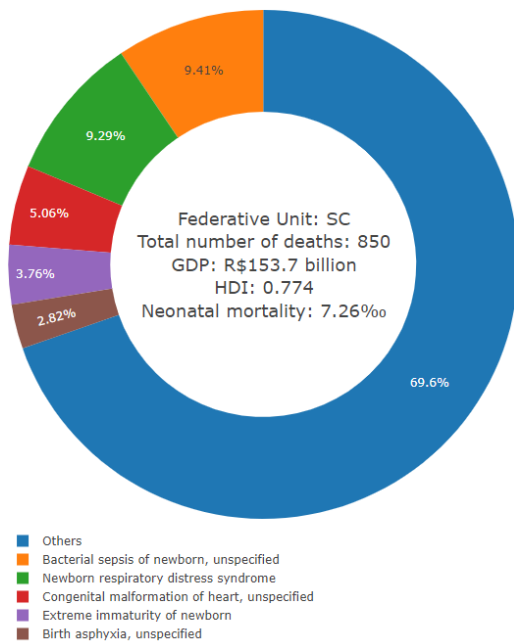

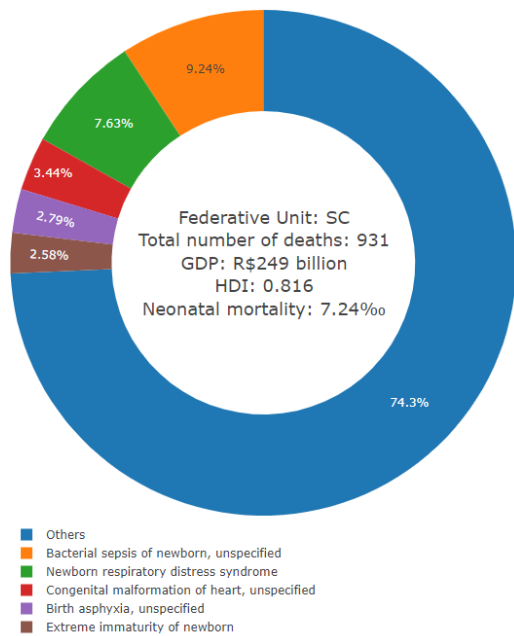**Figure 3: Relation between mortality, GDP and HDI in Santa Catarina in 2010**

XI Computer on the Beach
September 2 to 4, *Baln. Camboriú, SC, Brazil*

Pimentel et al.

Figure 4: Relation between mortality, GDP and HDI in Santa Catarina in 2015
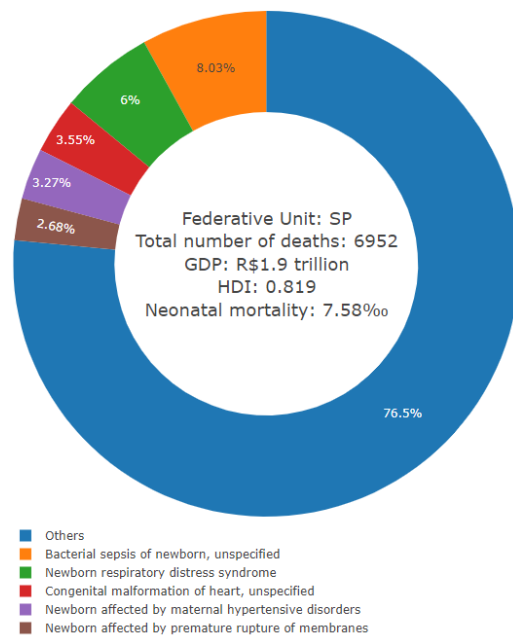

Figure 5: Relation between mortality, GDP and HDI in São Paulo in 2010
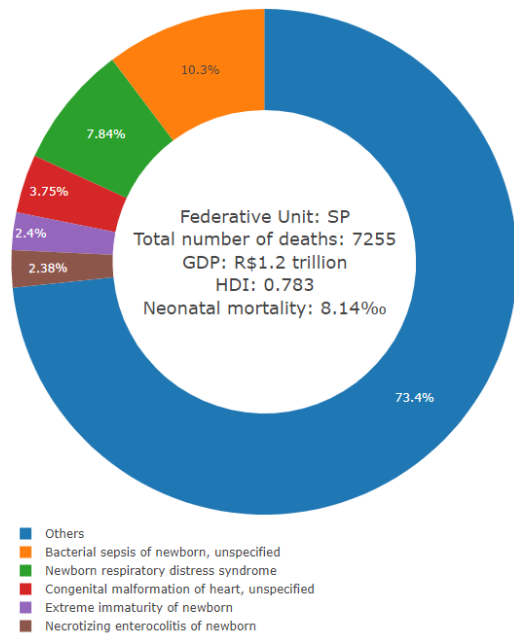

Figure 6: Relation between mortality, GDP and HDI in São Paulo in 2015

The donut charts show that Santa Catarina's GDP and HDI increased considerably, however the neonatal mortality rate decreased 0.02 per thousand. In São Paulo, the same happens for GDP and HDI, but neonatal mortality decreased by 0.66. There were also cases where neonatal mortality increased reasonably, such as Roraima (RR), starting with 7.50 per thousand in 2010 and ending up with 10.52 per thousand in 2015. Something similar happened in Maranhão (MA), Rio Grande do Norte (RN) and Sergipe (SE), but the neonatal mortality rate did not increase too much.

3.3 Relation between first trimester of prenatal care, immunizations of pregnant women and mortality

The third and last visualization proposed in this work depicts the relation between the proportion of pregnant of women who started prenatal care in the first trimester of pregnancy, immunizations of pregnant women and neonatal mortality. This graph in particular refers to health areas in Brazil, containing 438 entries, one for each area. For this visualization, a scatter plot with a colorbar mortality indicator, where red means high and blue means low, were chosen. The horizontal axis represents the proportion of vaccinated pregnant women and the vertical axis represents the proportion of pregnant women in the first trimester of prenatal care. The expected result was a graph with two clusters, where one of them would be full of red dots with overall low percentage of vaccinated pregnant women and low percentage of pregnant women who have started prenatal care in the first trimester of pregnancy, and the other one would be full of blue dots with overall high percentages of the same indicators. The actual result, depicted in figures 7 and 8, does not look like what was expected. There is only one region resembling

XI Computer on the Beach
September 2 to 4, 2020, Baln. Camboriú, SC, Brazil

Pimentel et al.

a cluster, full of blue, red and purple dots, the latter representing a not too high and not too low neonatal mortality. However, it is possible to notice that the region fits a linear regression line decently, specially in the years 2006 and 2010, as seen in figures 8 and 9, respectively.
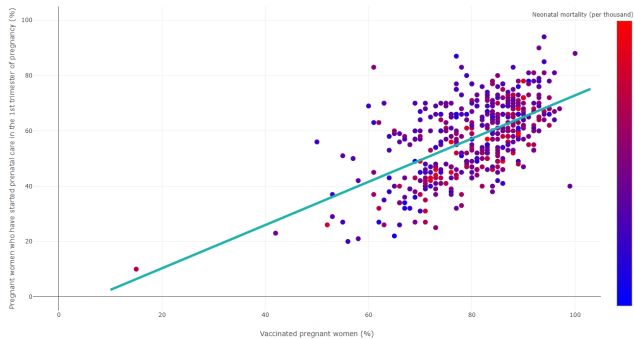

Figure 7: Relation between first trimester of prenatal care, immunizations of pregnant women and mortality in 2000
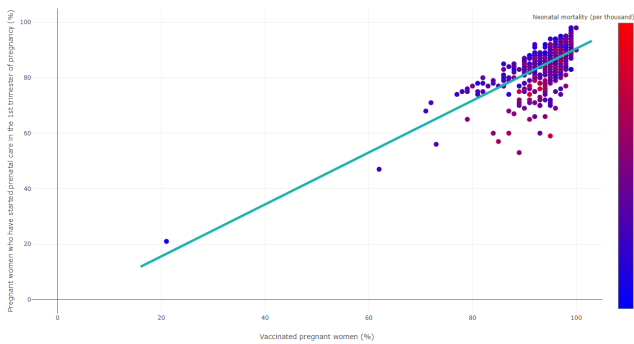

Figure 8: Relation between first trimester of prenatal care, immunizations of pregnant women and mortality in 2015
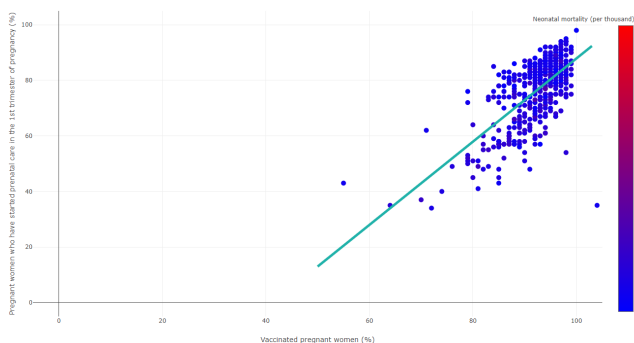

Figure 9: Relation between first trimester of prenatal care, immunizations of pregnant women and mortality in 2006

Using the linear regression concept, it is possible to estimate not only future neonatal mortality, but many other indicators as well, such as the proportion of vaccinated pregnant women and proportion of women in the first trimester of prenatal care. This, however, requires the usage of certain machine learning tools, such as Keras [7] and TensorFlow [8], not covered in this work.

4 CONCLUSION

Being able to visualize the data related to neonatal mortality is of paramount importance to mitigate the problem, as it provides an information rich way to understand its causes, allowing further studies to propose applicable solutions for it, such as the use of deep learning tools to estimate the amount of women that will start their prenatal care in the first trimester of pregnancy based on their vaccination history, and intervene if possible. It is expected that this work, open source on GitHub (github.com/lucasrdrgs/VisualizacoesCOTB), is useful for such understanding and mitigation of the problem.

ACKNOWLEDGEMENTS

REFERENCES

[1] Ministério da Saúde (Brazil). Datasus. URL http://datasus.saude.gov.br/.
[2] Governo Federal and Ministério da Economia (Brazil). Ibge. URL https://ibge.gov.br/.
[3] Python Organization. Python. URL https://www.python.org/.
[4] Wes McKinney. Pandas – python data analysis library. URL https://pandas.pydata. org/.
[5] Travis Oliphant. Numpy. URL https://numpy.org/.
[6] Plotly. Dash by plotly. URL https://plot.ly/dash/.
[7] François Chollet. Keras: The python deep learning library. URL https://keras.io/.
[8] Google Brain Team. Tensorflow: An end-to-end open source machine learning platform. URL https://www.tensorflow.org/.