Proceedings of Statistics Canada Symposium 2016
Growth in Statistical Information: Challenges and Benefits

Epidemiological observatory on Brazilian health data

Raphael de Freitas Saldanha[1], Ronaldo Rocha Bastos[2]

Abstract

The Unified Brazilian Health System (SUS) was created in 1988 and, with the aim of organizing the health information systems and databases already in use, a unified databank (DataSUS) was created in 1991. DataSUS files are freely available via Internet. Access and visualization of such data is done through a limited number of customized tables and simple diagrams, which do not entirely meet the needs of health managers and other users for a flexible and easy-to-use tool that can tackle different aspects of health which are relevant to their purposes of knowledge-seeking and decision-making. We propose the interactive monthly generation of synthetic epidemiological reports, which are not only easily accessible but also easy to interpret and understand. Emphasis is put on data visualization through more informative diagrams and maps.

Key Words: Health data dissemination; dynamic algorithms; data on the web.

1. Introduction

The production and availability of health data is crucial to the improvement of health systems, guiding the process of decision-making and research (Handley et al. 2015). The availability of health data in developing countries are often not adequate (Boerma and Stansfield 2007), although Brazil has been investing on this field since the 90's. With the re-democratization process that started on 1988, the new Brazilian constitution established a Unified Health System, and in 1991 the Department of Informatics of the Brazilian Health System (DataSUS) was created.

Since its creation, DataSUS has been collecting information on mortality, morbidity, service provision and health resources, providing data for decision making and research. Currently, all data collected by DataSUS is available on the department's website (http://datasus.saude.gov.br).

One major concern about health data in developing countries, besides the production of data itself, is the use of data that is already produced but not yet used, leading to the need to strengthen the analysis of the data (Boerma and Stansfield 2007). This problem is also faced in Brazil: following its own purpose, DataSUS is more focused on data production and dissemination than on analysis and this is reflected on its website. Data access requires some training and knowledge of epidemiology as only raw files or tables can be produced, requiring from the user a second step of analysis to transform the data into information.

Data from 2014 shows that in Brazil, 85% of public health units have access to the Internet and 99% of M.D.s had accessed the Internet in the previous 12 months (Comitê Gestor da Internet no Brasil 2015), but the statistics on the DataSUS webpage use shows an average of 210,000 access throughout the same year (DataSUS 2015). Considering that Brazil has more than 200,000 administrative professionals in the health system (MINISTÉRIO DA SAÚDE 2016), the number of accesses to the DataSUS website is far away from expected.

Following the efforts of collecting data, more attention must be paid to information communication of health data and indicators. The communication of health data aimed at decision-makers should go beyond crude indicators and raw numbers, emphasizing the relationship of present data to past trends and others indicators, establishing relations

1 UFJF – Programa de Pós-Graduação em Saúde Coletiva, Rua José Lourenço Kelmer, s/n. Campus Universitário – São Pedro – Juiz de Fora – Brazil. ZIP Code: 36036-900 (rfsaldanha@gmail.com);
2 UFJF – Departamento de Estatística, Rua José Lourenço Kelmer, s/n. Campus Universitário – São Pedro – Juiz de Fora – Brazil. ZIP Code: 36036-900 (ronaldo.bastos@ufjf.edu.br).

between subgroups and geographic areas and relying on charts and maps to help deliver information (AbouZahr, Adjei, and Kanchanachitra 2007).

The current project aims to contribute to the needs of health managers, researchers and the population in general for a flexible and easy-to-use tool that can tackle different aspects of health which are relevant to their purposes of knowledge-seeking and decision-making. For this, we propose the interactive monthly generation of synthetic epidemiological reports, which are not only easily accessible but also easy to interpret and understand. Emphasis is put on data visualization through more informative diagrams and maps, as presented in the following section.

2. Methodology

The methodology of this project regards the data acquisition, processing and delivering of reports on a monthly basis, as shown below.

2.1 Data

The Brazilian Ministry of Health through DataSUS make aggregate and raw data available in their website. We use the raw data available in DBC format to create reports using the desired levels of aggregation. In these raw data files, each record presents information of one occurrence, such as a hospital admission for example.

The DBC files are downloaded on a monthly basis, converted to DBF and imported to the R language for statistical computing environment (R Core Team 2014). After reading the raw data, the databases are prepared and cleaned by dropping variables of no interest, converting units and labelling levels of categorical variables, following the documentation files that are also available in the DataSUS website. All steps are executed in an automatized way, creating ready-to-use data sets separated by categories of interest.

These categories of interest separate the records by the International Code of Diseases (ICD) informed in the diagnosis fields, following the same 22 aggregation criteria used by the Global Burden of Disease project (Department of Health Statistics and Information Systems WHO 2013).

Currently, we have been using the DBC files regarding hospital admissions, but other dimensions like ambulatory attendance and mortality are also available.

2.2 Processing

This step is dedicated to create reports for each category of interest and levels of aggregation. For this, four algorithms were developed regarding national, state, macro and micro regional levels for Brazil.

The creation of *html* reports is done using the R library *knitr* (Yihui 2015) and *markdown* (Allaire et al. 2014), and this is the most time demanding step of the process. Whenever possible, multi-thread strategies are applied using the *doParallel* package (Revolution Analytics and Weston 2014).

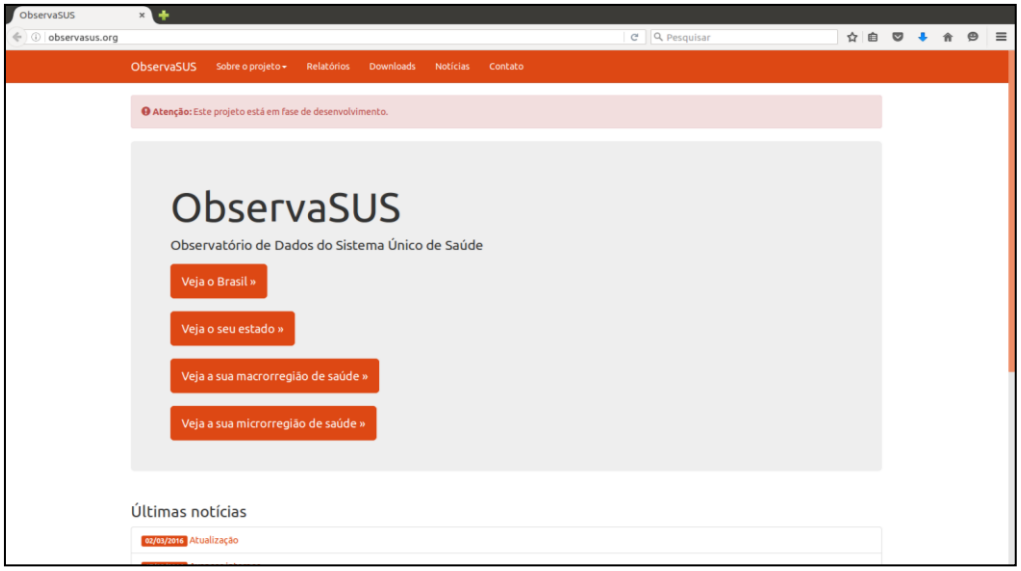
All algorithms we use are managed by a *subversion* repository, which enables fast and reliable code versioning, management and recovery for the developers.

2.3 Webpage

The reports created are made available in the project web-page (http://observasus.org). The architecture of the website was designed aiming at a simple and direct interface, requiring from the user a minimum level of knowledge on epidemiology.

The reports can be accessed directly from the first page (see Figure 2.2-1). After selecting the level of aggregation, the user is asked to choose the category of interest to be seen throughout the generated report, requiring from two to four clicks to get a report on the screen.

Figure 2.2-1
ObservaSUS main page



The levels of aggregation for macro and micro regions are accompanied by a link that helps the user to select the desired region. The same strategy is applied for the categories of interest, where the user can inform the name of a disease and point at the chosen category.

The website contains other sections: (1) a blog used to inform the users about updates and improvements made; (2) one 'about the project' section, with general information, team, technologies and the methodology used, (3) downloads area where the users can download the data sets mentioned in the 2.1 section and (4) an issue tracking system where the users can send suggestions and report errors.

3. Results

The current project is able to provide informative reports to several categories of interest and aggregation levels, helping to fill the gap between data access and information production, not acting as a substitute to DataSUS, but being a different front-end to the data. A sample of (part of) a typical report is shown in Figure 3-1.

More advances are expected to be developed using other databases that are already available, like ambulatory attendances and mortality, in a way to deliver a more complete picture of the health status in the country. The consequential increase in the amount of information produced needs to be followed by improvements in the interface, requiring the minimum set of skills from the user to access and fully comprehend the reports. Thus, delivering health information through informative and updated reports, this project is confident to be a reliable source of information for decision makers and the Brazilian population alike.

Figure 3-1
ObservaSUS report example

Acknowledgements

This project receives the support of the Federal University of Juiz de Fora, through the Post-Graduation Program on Public Health, the Geo-referenced Information Lab – LINGE of the Department of Statistics and the Centre for Advisory, Training and Health Studies – NATES.

References

AbouZahr, Carla, Sam Adjei, and Churnrurtai Kanchanachitra. (2007). "From Data to Policy: Good Practices and Cautionary Tales." *Lancet* 369 (9566): 1039–46. doi:10.1016/S0140-6736(07)60463-2.

Allaire, JJ, Jeffrey Horner, Vicent Marti, and Natacha Porte. (2014). "Markdown: Markdown Rendering for R." R package version 0.7.4.

Boerma, J Ties, and Sally K Stansfield. (2007). "Health Statistics Now: Are We Making the Right Investments?" *Lancet* 369 (9563): 779–86. doi:10.1016/S0140-6736(07)60364-X.

Comitê Gestor da Internet no Brasil. (2015). "TIC Saúde 2014: Pesquisa Sobre O Uso Das Tecnologias de Informação E Comunicação Nos Estabelecimentos de Saúde Brasileiros". São Paulo: CETIC.BR.

DataSUS. (2015). "Estatísticas de Acesso Ao TabNET: 2014". Brasília.

Department of Health Statistics and Information Systems WHO. (2013). "WHO Methods and Data Sources for Global Burden of Disease Estimates 2000-2011". Global Health Estimates Technical Paper WHO/HIS/HSI/GHE/2013.4. Geneva. http://www.who.int/healthinfo/statistics/GlobalDALYmethods_2000_2011.pdf?ua=1.

Handley, Kathleen, Ties Boerma, Cesar Victora, and Timothy Grant Evans. (2015). "An Inflection Point for Country Health Data." The Lancet Global Health, no. 15: 437–38. doi:10.1016/S2214-109X(15)00067-4.

MINISTÉRIO DA SAÚDE. (2016). “DataSUS.” http://www2.datasus.gov.br.

R Core Team. (2014). “R: A Language and Environment for Statistical Computing”. Vienna: R Foundation for Statistical Computing. http://www.r-project.org/.

Revolution Analytics, and Steve Weston. (2014). “doParallel: Foreach Parallel Adaptor for the Parallel Package.” R package version 1.0.8.

Yihui, Xie. (2015). “Knitr: A General-Purpose Package for Dynamic Report Generation in R.” R package version 1.9.