



Conference Report

SalHUD—A Graphical Interface to Public Health Data in Puerto Rico [†]

Humberto G. Ortiz-Zuazaga ^{1,*}, Roberto Arce-Corretjer ¹, Juan M. Solá-Sloan ² and José G. Conde ³

Received: 14 August 2015; Accepted: 10 October 2015; Published: 22 December 2015

Academic Editors: Mark Edberg, Barbara E. Hayes, Valerie Montgomery Rice and Paul B. Tchounwou

¹ Department of Computer Science, University of Puerto Rico, Río Piedras Campus, San Juan 00936, Puerto Rico; racorretjer@gmail.com

² Department of Computer Science, University of Puerto Rico, Bayamón Campus, Bayamón 00959, Puerto Rico; juan.sola@upr.edu

³ School of Medicine, University of Puerto Rico, Medical Sciences Campus, San Juan 00936, Puerto Rico; jose.conde1@upr.edu

* Correspondence: humberto.ortiz@upr.edu; Tel.: +787-764-0000 (ext. 4758); Fax: +787-773-1717

† This paper is an extended version of our paper published in the 2014 Minority Health & Health Disparities Grantees' Conference (MHHDGC), National Harbor, MD, USA, 1–3 December 2014.

Abstract: Purpose: This paper describes SalHUD, a prototype web-based application for visualizing health data from Puerto Rico. Our initial focus was to provide interactive maps displaying years of potential life lost (YPLL). Methods: The public-use mortality file for year 2008 was downloaded from the Puerto Rico Institute of Statistics website. Data was processed with R, Python and EpiInfo to calculate years of potential life lost for the leading causes of death on each of the 78 municipalities in the island. Death records were classified according to ICD-10 codes. YPLL for each municipality was integrated into AtlasPR, a D3 Javascript map library. Additional Javascript, HTML and CSS programming was required to display maps as a web-based interface. Results: YPLL for all municipalities are displayed on a map of Puerto Rico for each of the ten leading causes of death and for all causes combined, so users may dynamically explore the impact of premature mortality. Discussion: This work is the first step in providing the general public in Puerto Rico with user-friendly, interactive, visual access to public health data that is usually published in numerical, text-based media.

Keywords: health data; visualization; Javascript

1. Introduction

Under the Health Data Initiative, the U.S. Department of Health and Human Services releases large amounts of health data for public use. The mission of the initiative is “to improve health, health care, and the delivery of human services by harnessing the power of data and fostering a culture of innovative uses of data in public and private sector institutions, communities, research groups and policy making arenas” [1]. Software developers use data to create applications that make health information increasingly useful for individuals, communities, service providers and policy-makers [1].

This paper describes SalHUD, a prototype web-based application for visualizing health data from Puerto Rico. Our initial focus was to provide interactive maps displaying years of potential life lost (YPLL) from leading causes of death by municipality. YPLL is a measure of premature mortality in the population under study. YPLL is not reported by the Department of Health in Puerto Rico as part of its annual statistics reports. The most recent publication about YPLL in Puerto Rico dates from

1992 [2]. In order to bring attention to this issue, we decided to choose YPLL in Puerto Rico as the first health indicator to be displayed in SalHUD.

2. Methods

The public-use Puerto Rico Basic Mortality file for year 2008, containing unidentified data from death certificates of 29,100 deaths, was downloaded from the Puerto Rico Institute of Statistics website [3]. Mortality files from Puerto Rico are generated by the Puerto Rico Department of Health following standards of the National Vital Statistics System, National Center of Health Statistics, Centers of Disease Control and Prevention. Deaths are recorded according to underlying cause of death codes from the International Classification of Diseases (ICD-10) [4]. Data are published as comma-separated value (.csv) files, one line per death record, with age at death coded into two columns (namely, type of age units such as minutes, hours, days, weeks, months or years; and number of units). Municipality of residence at time of death is coded with a three-digit number, and underlying cause of death is coded into four columns containing ICD-10 codes. See Figure 1 for a portion of one of the files.

Data was processed with Python [5] to extract the municipality of residence, the age in years at death, and categorize the cause of death into one of the 10 leading causes of death in Puerto Rico for each record. The python program then computes YPLL before the age 75 for each record, and prepares a table with the total YPLL for each leading cause of death, and for all causes combined, per municipality. YPLL was calculated by subtracting the age at death from 75 [6]. Starting in 1996, the National Center for Health Statistics (NCHS) has been presenting YPLL for persons under age 75 because the average life expectancy in the United States is over 75 years [7]. We decided to do the same in order to provide numbers that are comparable to those reported by NCHS. In addition, life expectancy in Puerto Rico for 2008 was 78 years [8].

	A	B	C	D	E	F	G	H	
1	staymunnumunit	staymununit	placeocurrence	stayplacenumunit	stayplaceunit	placeresidence	countryresidence	monthdeath	year
2	XX	X	98	0	0	232	1	1	8
3	XX	X	1	10	1	308	1	1	1
4		4	6	98	0	0	260	1	1
5	XX	X		2XX		2	216	1	1
6	XX	X		1	2	4	144	1	2
7	XX	X		2	12	3	40	1	2
8	XX	X		2	3	5	200	1	2
9	XX	X		22	45	3	52	1	3
10	XX	X		1	4	2	128	1	5
11		20	6	98	0	0	56	1	5
12		1	6	98	0	0	52	1	5
13	XX	X		6	2	2	292	1	3
14	XX	X		2XX		1	40	1	7
15	XX	X		3	12	3	116	1	6
16	XX	X		2XX		2	68	1	3
17	XX	X		3XX		1	68	1	3
18	XX	X		13	2	5	164	1	8
19	XX	X		12	2	5	164	1	8

Figure 1. A portion of the Puerto Rico Basic Mortality file for year 2008. The full file contains 29,100 records.

We constructed an R [9] prototype application to construct a choropleth, a map of Puerto Rico shaded by combined YPLL in each municipality, as a proof of concept. This choropleth was static, and displayed only the total YPLL for each municipality, not the YPLL for a particular cause of death. EpiInfo [10] was used to validate the python program and check the YPLL values we had calculated. The final interactive maps were developed with AtlasPR, a D3 Javascript map library [11]. Additional Javascript, Hyper Text Markup Language (HTML) and Cascading Style Sheet (CSS) programming was required to display maps as a web-based interface where users can select YPLL from all causes combined or one of the 10 leading causes of death, and have the map update dynamically.

3. Results and Discussion

YPLL for all municipalities are displayed on a map of Puerto Rico for each of the ten leading causes of death and for all causes combined, so users may dynamically explore the impact of premature mortality (see Figure 2). A drop-down menu provides a list of causes of death to select what YPLL map to view. Map segments representing municipal territories are shaded in a gradient based on a linear interpolation between minimum and maximum values of the distribution of YPLL among municipalities.

In addition, the actual number of YPLL for each municipality is integrated to each area, so that clicking on any segment of the map activates a pop-up window displaying the YPLL for the municipality. Figure 3 shows the pop-up dialog box.

A video of the application is available at <https://www.youtube.com/watch?v=YIXvXLQTZfw>. A working prototype is available at: <http://www.hpcf.upr.edu/~humberto/salHUD/>. The full source code for the application is available as well [12].

The resulting application displays a summary of the impact of premature death by the leading causes of death in Puerto Rico per municipality. This display is much more accessible to non-technical users than the original data files (See Figure 1). In addition, presenting the data in a graphical format is a major improvement over the usual tabular layout commonly used in vital statistics reports, minimizing information overload and optimizing extraction of information from data.

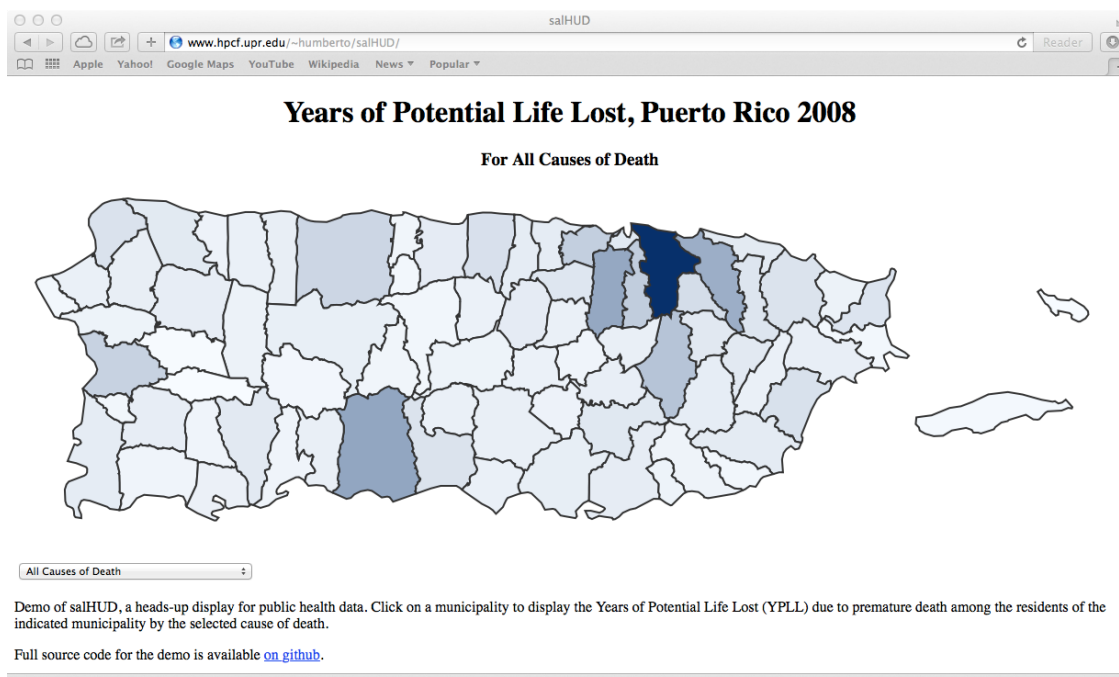


Figure 2. Screenshot of the SalHUD demo application shaded by all causes years of potential life lost (YPLL) in each municipality in Puerto Rico.



Figure 3. Screenshot of the SalHUD demo application showing the total YPLL for Chronic Lower Respiratory Airway Disease in the municipality of Ponce in 2008.

One key component of the Health Data Initiative is to foster a culture of innovative uses of data at the community level [1]. This can only be achieved by presenting data in a manner that is easy to understand and that engages citizens from all sectors and ranges of society in the analysis of data for the common good. SalHUD is a step in that direction.

4. Conclusions

SalHUD displays YPLL distribution among municipalities in Puerto Rico in an interactive, graphical environment, stimulating exploration and engagement by the user. This is the first step in providing the general public in Puerto Rico with user-friendly, dynamic access to public health data that is usually published in numerical, text-based media. A similar approach can be used to visualize other types of health data and to visualize health indicators in other communities.

Acknowledgments: Humberto G. Ortiz-Zuazaga is partially supported by INBRE grant P20GM103475 from the National Institute of General Medical Sciences, and José G. Conde by RCMI grant G12MD007600 from the National Institute on Minority Health and Health Disparities, National Institutes of Health. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions: Humberto G. Ortiz-Zuazaga developed the python script that computes overall and by municipality YPLL from the mortality data. Roberto Arce-Corretjer developed the javascript code to integrate the YPLL data into AtlasPR and display the maps. Juan M. Solá-Sloan developed the HTML/Javascript to select the causes of death. José G. Conde developed the concept, wrote most of the article, assisted in the computation of YPLL and provided biomedical expertise.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Health Data—HHS Idea Lab. Available online: <http://www.hhs.gov/idealab/what-we-do/health-data/> (accessed on 12 September 2015).
2. Ramirez de Arellano, A.B. The death divide: Differentials in premature mortality by gender in Puerto Rico. *Bol. Asoc. Med. P. R.* **1992**, *84*, 11–14. [PubMed]
3. Instituto de Estadísticas de Puerto Rico. Basic Mortality. 2008. Available online: http://www.estadisticas.gobierno.pr/iepr/Estadisticas/Basesdedatos/Salud.aspx#CDC_mort_ba (accessed on 12 August 2015).
4. International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10). Available online: <http://apps.who.int/classifications/icd10/browse/2015/en> (accessed on 12 September 2015).

5. Python Software Foundation. Python Language Reference, Version 2.7. Available online: <http://www.python.org/> (accessed on 12 September 2015).
6. Injury Prevention & Control: Data & Statistics (WISQARSTM): Atlanta: Centers for Disease Control and Prevention (U.S.), 5.3 Definitions for Years of Potential Life Lost. Available online: http://www.cdc.gov/injury/wisqars/fatal_help/definitions_ypll.html (accessed on 12 September 2015).
7. National Center for Health Statistics. *Health, United States, 2014: With Special Feature on Adults Aged 55–64*; National Center for Health Statistics: Hyattsville, MD, USA, 2014.
8. Informe de la Salud de Puerto Rico. 2014. Available online: [http://www.salud.gov.pr/Estadisticas-Registros-y-Publicaciones/Estadisticas Vitales/Informe de la Salud en PUerto Rico 2014.pdf](http://www.salud.gov.pr/Estadisticas-Registros-y-Publicaciones/Estadisticas-Vitales/Informe-de-la-Salud-en-PUerto-Rico-2014.pdf) (accessed on 12 September 2015).
9. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015; Available online: <https://www.R-project.org> (accessed on 13 August 2015).
10. Dean, A.G.; Arner, T.G.; Sunki, G.G.; Friedman, R.; Lantinga, M.; Sangam, S.; Zubieta, J.C.; Sullivan, K.M.; Brendel, K.A.; Gao, Z.; et al. *Epi Info, a Database and Statistics Program for Public Health Professionals*; CDC: Atlanta, GA, USA, 2011.
11. Atlas PR. Available online: <http://miguelrios.github.io/atlaspr/> (accessed on 12 September 2015).
12. SalHUD Source Code. Available online: <https://github.com/humberto-ortiz/salHUD> (accessed on 19 September 2015).



© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).