Data Visualization with IBM Watson Analytics for Global Cancer Trends Comparison from World Health Organization

Kelvin K. F. Tsoi, Stanley Ho Big Data Decision Analytics Research Centre and Jockey Club School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China

Felix C. H. Chan, Stanley Ho Big Data Decision Analytics Research Centre, Chinese University of Hong Kong, Hong Kong, China

Hoyee W. Hirai, Stanley Ho Big Data Decision Analytics Research Centre, Chinese University of Hong Kong, Hong Kong, China

Gary K. S. Keung, Stanley Ho Big Data Decision Analytics Research Centre, Chinese University of Hong Kong, Hong Kong, China

Yong-Hong Kuo, Stanley Ho Big Data Decision Analytics Research Centre, Chinese University of Hong Kong, Hong Kong, China

Samson Tai, IBM China/Hong Kong Limited, Hong Kong, China

Helen M. L. Meng, Stanley Ho Big Data Decision Analytics Research Centre and Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong, Hong Kong, China

ABSTRACT

Visual analytics is widely used to explore data patterns and trends. This work leverages cancer data collected by World Health Organization (WHO) across a hundred of cancer registries worldwide. In this study, the authors present a visual analytics platform, IBM Watson Analytics, to explore the patterns of global cancer incidence. They included 26 forms of cancers from eight different geographic regions which are United States, the United Kingdom, Costa Rica, Sweden, Croatia, Japan, Hong Kong and China (Shanghai). An interactive interface was applied to plot a choropleth map to show global cancer distribution, and line charts to demonstrate historical cancer trends over 29 years. Subgroup analyses were conducted for different age groups. With real-time interactive features, one can easily explore the data with a selection of any cancer type, gender, age group, or geographical region. This platform is running on the cloud, so it can handle data in huge volumes, and is accessible by any computer connected to the Internet. IBM Watson Analytics released a latest version named "IBM Watson Analytics New User Experience" in the end of 2016. The new version streamlined the process to add data, discover data meaning and display result visually. The authors discuss the new features in the end of this paper.

KEYWORDS

Cancer Incidence, Choropleth Map, Global Comparison, Interactive Interface, Visual Analytics

1. INTRODUCTION

Visual analytics have been shown to be effective for data exploration (Keim, 2002), but the requirement of computational power is high for global comparisons of disease trends. Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time (Mell & Grance, 2011). The elasticity of cloud computing meets the demand of high computational power for data exploration. Therefore, cloud computing is one of major enablers for explorations of data. This is a general shift of computer processing, storage, and software delivery away from the traditional desktop computers and local servers towards the cloud (Saranya & Sunitha, 2012).

Cancer is one of the leading causes of morbidity and mortality worldwide, with approximately 14 million new cases and 8.2 million cancer related deaths in 2012. It is expected that annual cancer cases will rise to 22 million within the next 2 decades. The prevalence of cancer varies by gender, age, ethnicity, geographical location, economic status, and so on. Generally, the cancer causes of death were common on breast, lung, liver, stomach, colon and rectum (International Agency for Research on Cancer World Health Organization, 2014). Although the age-standardized incidence rates on some cancer showed stable trends, but the prevalence of cancer has grown along with the ageing population. To better understand the progression of cancer, cancer registries were set up in different countries. The first population-based cancer registry was in Germany Hamburg in 1926 (Wagner, 1991). Cancer registries have been widely used in epidemiological research, so the World Health Organization (WHO) formed an International Agency for Research on Cancer (IARC) to collect cancer registry data across different countries. Descriptive studies use the registry database to examine differences in the incidence of cancer for different patient characteristics (Parkin, 2006). The data volume of the global cancer incidence is huge, so visual analytics can help to enhance the data interpretation on disease distribution and trends. It simplifies complex data to intuitive and interactive visual representation for faster and better understanding of the meaning of the data.

The main contribution in this study is to use a visual analytics platform (IBM Watson Analytics) to aid in visualizing differences in cancer trends and patterns embedded within data sourced from WHO cancer registries. The research aims to answer several major questions presented to us from cancer researchers, including, but not limited to (i) What are the top-ranking forms of cancers (ii) across different regions, (iii) across gender, (iv) over the years, (v) across high- and mid-income regions, and (vi) across different age groups? In the following session, we provide a visualization of the WHO data that enables us to intuitively answer these questions. In the future, such intuition can guide us in further explorations. For example, we can use the patterns to show the effectiveness of colorectal cancer (CRC) screening programme in US, and compare the results with some regions that do not have screening programme.

We selected the Choropleth map and the traditional line charts for demonstration. The Choropleth map used to present the population density in different geographical regions (MacEachren, Brewer, & Pickle, 1998). In this study, the regions with high volume of cancer incidence will appear with darker colors on the map. The line chart is the traditional way of presentation which shows data trends along timeline. In this study, line charts are used to demonstrate cancer trends across regions and different population groups, such as for different gender. A matrix of line charts is also developed for cross-sectional comparison between different age groups across the regions.

The structure of this paper is organized as follows. Section 2 discusses related work in the academic field. Section 3 focuses on data structure of cancer registry from WHO, while Section 4 highlights data categorization for subgroup comparison. Section 5 provides details on the visual analytics environment, while Section 6 presents the application scenario. Section 7 discusses some new features in IBM Watson Analytics New User Experience, the latest release of IBM Watson Analytics. Finally, Section 8 provides a conclusion with thoughts for future research.

Related Content

Patient Safety Concerns among Emergency Medical Staff and Patients
Pi-Fang Hsu, Wen-Chun Tsai and Chia-Wen Tsai (2013). *International Journal of Privacy and Health Information Management (pp. 29-52).*
www.igi-global.com/article/patient-safety-concerns-among-emergency/77005?camid=4v1a

Visual Data Mining in Physiotherapy Using Self-Organizing Maps: A New Approximation to the Data Analysis
Yasser Alakhdar, José M. Martínez-Martínez, Josep Guimerà-Tomás, Pablo Escandell-Montero, Josep Benitez and Emilio Soria-Olivas (2012). *Medical Applications of Intelligent Data Analysis: Research Advancements (pp. 187-194).*
www.igi-global.com/chapter/visual-data-mining-physiotherapy-using/67258?camid=4v1a

Mathematical Programming and Heuristics for Patient Scheduling in Hospitals: A Survey

Daniel Gartner and Rema Padman (2018). *Health Care Delivery and Clinical Science: Concepts, Methodologies, Tools, and Applications (pp. 420-439).*

www.igi-global.com/chapter/mathematical-programming-and-heuristics-for-patient-scheduling-in-hospitals/192684?camid=4v1a

Personalized Mobile Applications for Remote Monitoring

Miguel A. Laguna and Javier Finat (2013). *International Journal of E-Health and Medical Communications (pp. 1-11).*

www.igi-global.com/article/personalized-mobile-applications-remote-monitoring/77301?camid=4v1a