Smoke Till My Last Breath Is So Good: An Interactive Visual Analytics Dashboard for Exploring and Analyzing Smokingn Data

Liu Jinlong, Lyu Cheng, Ji Xinyi

Abstract — The way we manage, analyze and leverage data has changed in many industries. One of the most likely areas where it can be applied to make a change is healthcare. By leveraging on visualized data, analysts and researchers can catch the trends and perspectives easily and efficiently. In fact, related institutions have to collect and manage mass of data to do healthcare and sociology studies. In sight of this need, we sought to make smoking data as meaningful and clear as possible. We designed a dynamic and interactive visual analytics dashboard to allow analysts and researchers to explore and analyze data from International Mortality and Smoking Statistics (IMASS), Institute of Health Metrics and Evaluation (IHME) and Global Burden of Disease (GBD) to gain valuable insights about worldwide prevalence of smoking, consumption of tobacco products and death caused by smoking-related diseases.

1 Introduction

Tobacco use is a major preventable cause of premature death and disease worldwide. Currently, approximately 5.4 million people die each year due to tobacco-related illnesses[1]. Smoking increases human beings' risk of developing more than 50 serious health conditions. Some diseases may be fatal, and others can cause irreversible long-term damage to your health. People can become ill: if you smoke yourself or if people around you smoke (passive smoking)[2].

The big data revolution provides vast advanced opportunities both for business purposes and academic studies. IMASS(International Mortality And Smoking Statistics)[3] collected updated Population and Mortality data of most countries and maintains a comprehensive Excel database system with smoking related data such as Consumption of tobacco products, Prevalence of smoking, Mortality from the major smoking-related diseases. Meanwhile Institute of Health Metrics and Evaluation (IHME), Global Burden of Disease (GBD) [4]manage the statistics on Death rates and absolute number of premature deaths from smoking and secondhand-smoke across all regions and countries. We merge the smoking-related data from the these data source.

This paper report covers our research and development efforts to design and implement a web-based visual analytics tool for supporting the analysis and visualisation of smoking data. Section 1 provides a general introduction of our project. Following which, we will cover our motivations and objectives of this project in Section 2. Section 3 provides details of our visualization of our dashboards. Section 4 involved the technology we used for the application. Section 5 will highlight some insights we obtained from our project, demonstrating how a user can interact with the dashboard to uncover useful information.

2 MOTIVATION AND OBJECTIVES

Our aim is to show perspectives of smoking and correlates, determinants and consequences about smoking with visual analytics. Meanwhile reveal the trends of smoking and other perniciousness attributed to smoking of specific area or countries, while compared to worldwide value, to bring to attentions for both individuals and countries.

3 THE VISUALIZATIONS

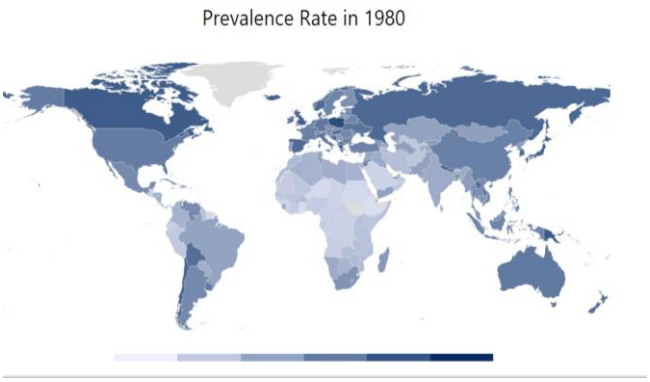
3.1 WORLD MAP

As we have collected several data sets with related smoking attribute, we need to process them and visualize them to the audience for their better understanding of current smoking issue among the world. Row data set is huge and not understandable by just reading it, by analysing the structure and the meaning of the data, our group came up with a visualization solution to better present the large data set.

3.1.1 heat map

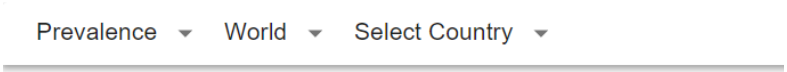
As our data is based on Topology, we decided to use Heatmap to visualize the data for each topic/year and country. In this chart, D3 is used and all countries are assigned by ISO country code. Darker color represent higher rates/indexes and on hover to every country, data details will be shown. Legend contains the range of data i.e. 20%-30%, on hover to each section, those counties with data fall in this range will be highlighted on the map. For each country on the map, it will pop up a model by clicking which contains the detail breakdown of that country with selected cluster and year.

Prevalence Rate of smoking over the world:



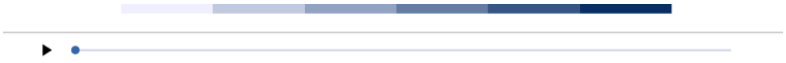
3.1.2 Type and area selection

Data type option is available via the drop-down selection box, by changing the data type and year, the map will change accordingly. Zoom in function is also available via the area drop down selection box, by selecting the area, the map focus will change accordingly.



3.1.3 Year bar

Below the map, a time bar is available, by sliding the time bar, the data will change the selected year and a auto play option is provided.

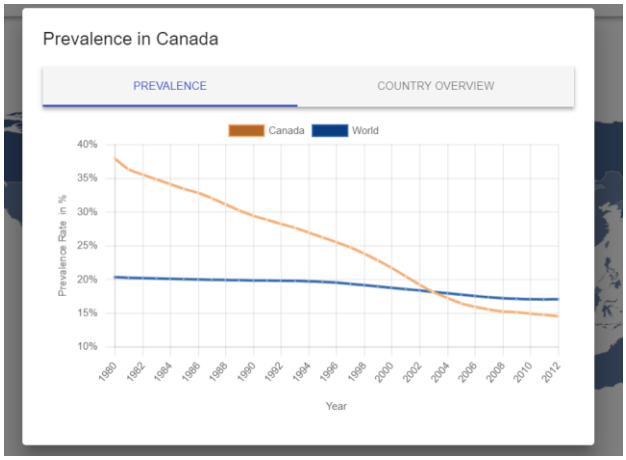


3.1.4 LINE CHART

When user choose to see the breakdown of certain country, a line chart of that country with selected data type will be shown. It plots the data in the line

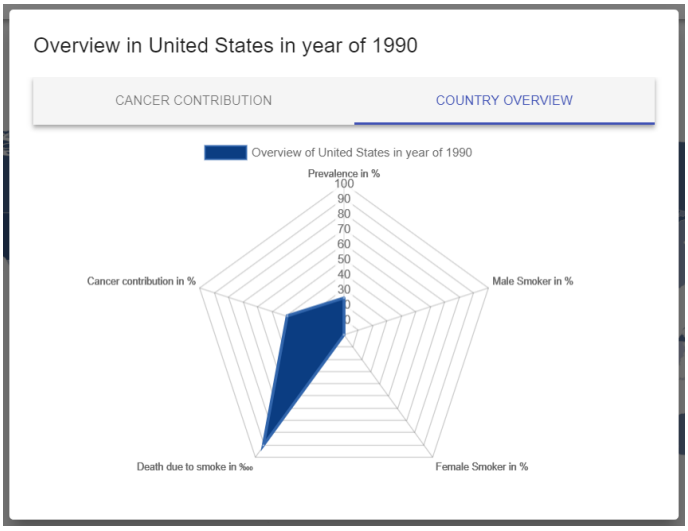
chart with year as x axis. Together, the world average data will also be shown in the same graph to make a contrast with the country, users can see the comparison in between. On hover to each data point, the data will be shown on the tooltips and ranking of that country in that year will also be provided.

Prevalence in one selected country:



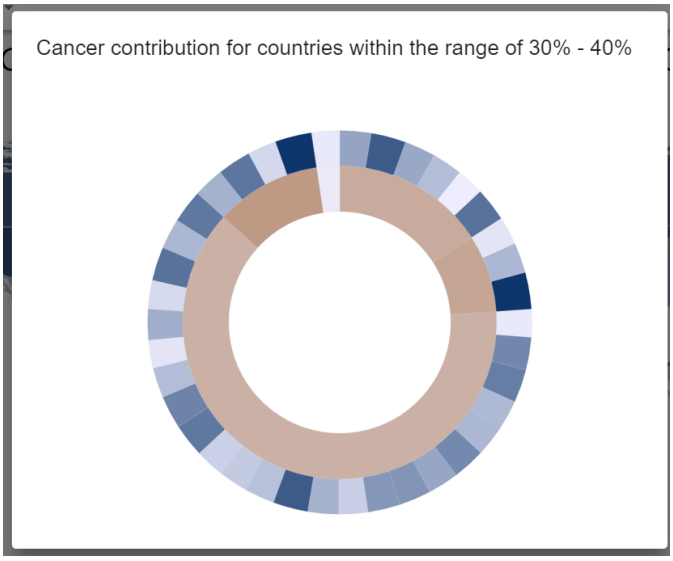
3.1.5 Radar chart

As we collected 5 data set with percentage as data, we will also provide the aggregated radar chart to user by clicking the country on the map. This radar chart will plot different data types in the same graph, since some types of the data might not be available in certain years, the rader chart might be incomplete with some data lacking. If the data is not available in that year/country, the data point will remain as 0 and showing notification when user hover on each data point.



3.1.6 Sunbursts

We use Sunbursts to visualize the smoking data in each area and selected range(Via legend). Heatmap legend are divided by data range, by clicking on each section, the sunburst graph will be given. The Sunbursts graph has three levels, Legend Range(Root) -> continent -> country. Darker color represent higher data. When hover to each section, the data of that section will be given, for continent it will provide the average in that range. This graph is meant to help user to zoom in to each continent and make comparison. If user is more interested to certain area and certain range of countries, they can have a clearer view through this graph. Size and color of each section all represent the data figure among the rest.

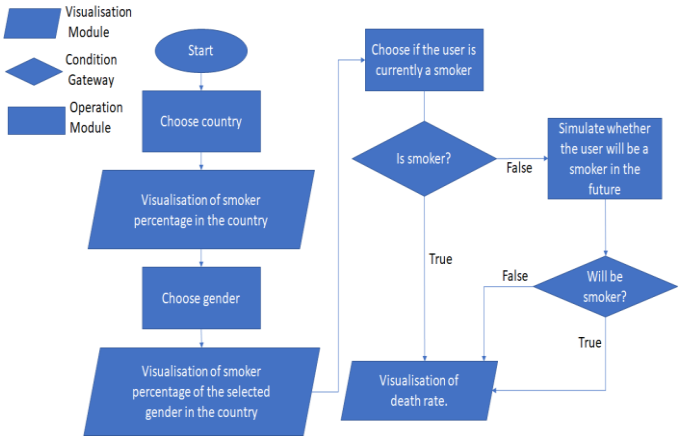


3.2 Fate

As the name tells, this subsection of the visualisation project aims to demonstrate the death rate due to smoking of any specific user of the application through a step-by-step procedure. In the interest of health, our application aims to discourage people from smoking, however, the mere global data cannot give a direct perception to each individual. Hence, we plan to visualise the the death rate due to smoking-related diseases of a user. Seeing the final fate of oneself, he or she may be demotivated from smoking.

3.2.1 Workflow

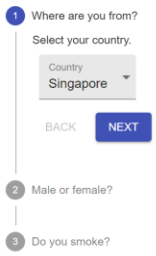
To visualise the death rate due to smoking, the personal information of the user is required. According to the data structure of our project, the country, gender are the compulsory parameters needed to find the death rate of a smoker. Hence the following workflow was designed.

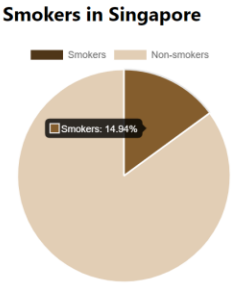


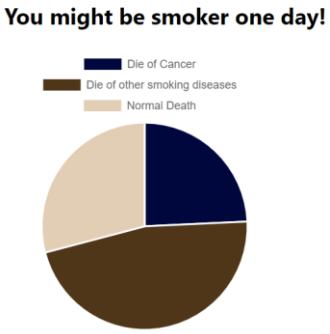
Notice that after the visualisation of a specific gender in the selected country, there appears a “Choose if the user is currently a smoker” and a condition gateway. This is simply because there is no way that a non-smoker will die due to smoking. Hence, we only display the result for the current smokers. However, there is no guarantee that the user will not become a smoker some day in the future. For the non-smokers, we process a random event to simulate whether he/she will be a smoker(The possibility is according to the smoker percentage in the selected country and of the selected gender). If the user so lucky that he/she will not become a smoker in the entire life, the displayed result will be zero per cent chance of smoking-related death.

We have designed an interactive visualization so that the pie chart on the right-hand side displaying smoker percentage of death rate is synchronised with the user’s selection in the stepper on the left-hand side. We break down the whole process into three steps as follows:

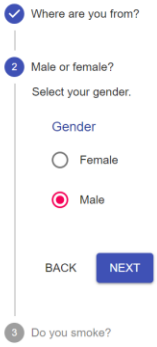
3.2.2 Country Selection

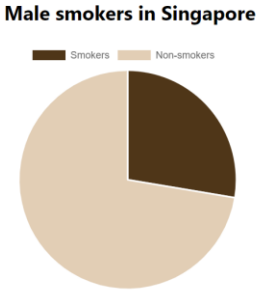




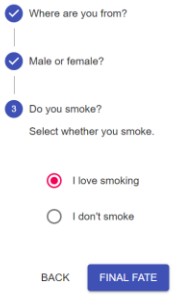


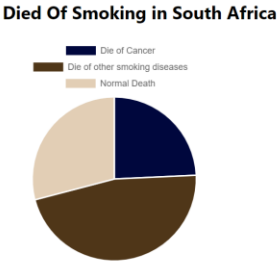
3.2.3 Gender Selection

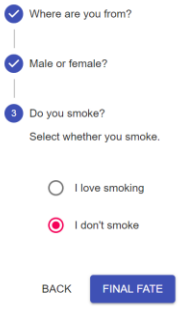


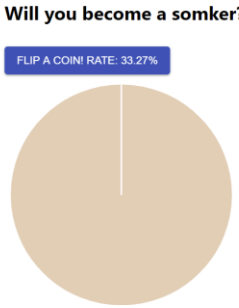


3.2.4 Smoker Selection









If “I don’t smoke” is selected, the user can press the button “FLIP A COIN” to simulate whether he/she become a smoker one day in the future.

4 The Application

The frontend of this project is developed based on the Reactjs framework. React makes it painless to create interactive UIs and allows to build encapsulated components that manage their own state, then compose them to make complex UIs[5]. There are four core component libraries we used in the project.

Material-ui: React components that implement Google's Material Design.[6]

Charjs: Simple, clean and engaging charts for designers and developers.[7]

Reactvis.js: React-vis is a React visualization library. It's been designed with the principles of React-friendly, High-level and customizable, Expressive and Industry-strong.[8]

D3.js: D3.js is a JavaScript library for manipulating documents based on data.[9]

The backend is supported by a Java servlet based on spring framework.

Both frontend and backend are deployed on a Google cloud server instance.

5 Insights

1.*The smoking rate is decreasing over the world.* From 1980 to 2012, the global smoking rate decreased except a fraction of countries. People gain more healthcare knowledge and pay more attention to their health.

2.*Smoking rate of both male and female in Europe is higher than that of countries from other continents.*

There may be some correlation between a country's economic/social conditions and smoking to make this special apperance. To further explore, we can find the deep relation of that.

6 CONCLUSION

The demonstration of our dashboard's potential using smoking data shows its ability to enable exploration of a large dataset and allow analysts to identify and catch information easily.

Besides the prevalence, this dashboard allow users to view the diseases rating because of smoking to brought health to the forefront.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Kam Tin Seong of Singapore Management University for his kind support and guidance in the process of this project.

REFERENCAES

[1]CDC – Centers for Disease Control and Prevention. (n.d.). Retrieved from https://www.cdc.gov/tobacco/global/index.htm

[2]NHS - Common-health-questions. (n.d.). Retrieved from https://www.nhs.uk/common-health-questions/lifestyle/what-are-the-health-risks-of-smoking/

[3] IMASS - International Mortality And Smoking Statistics. Retrieved from http://www.pnlee.co.uk/imass.htm

[4] Institute of Health Metrics and Evaluation (IHME), Global Burden of Disease (GBD). Retrieved from http://ghdx.healthdata.org/gbd-results-tool

[5]React – A JavaScript library for building user interfaces. (n.d.). Retrieved from https://reactjs.org/

[6]The world's most popular React UI framework - Material-UI. (n.d.). Retrieved from https://material-ui.com/

[7]Chartjs. (n.d.). Chart.js. Retrieved from https://www.chartjs.org/docs/latest/

[8](n.d.). Retrieved from https://uber.github.io/react-vis/

[9]Bostock, M. (n.d.). Data-Driven Documents. Retrieved from https://d3js.org/