A Story Behind GDP

CSE 591 Computer Visualization Portfolio

Ersi Zha , Master student, Arizona state university

Abstract—This paper introduces a system that focuses on a story behind GDP by using several visualization methods such as line charts, choropleth maps and bubble plots. The main technology we used to implement this project is D3.js. This project visualizes datasets in five different aspects: GDP, Urban Development, Education, Health and Infrastructure. The system provides users many ways to explore the facts behind GDP interactively.This paper presents the design ideas, the implementation process, the results of this system and my contribution to this project.

Key Words—D3.js, GDP, story, visualization

1 INTRODUCTION

The topic of this project is to explore how GDP impacts the standard of living of a country and make a comparison between other countries in the world. GDP stands for gross domestic product and is defined by the OECD (Organization for Economic Cooperation and Development) as "an aggregate measure of production equal to the sum of the gross values added of all resident institutional units engaged in production"[1]. GDP is commonly considered to be a measure of the standard of living of a country. But sometimes GDP cannot comprehensively and accurately reflect the standard of living of a country because it depends on many aspects, such as rural and urban development, health status, poverty and infrastructure.

Nowadays, many people are still confused about the relationship between GDP and these aspects. Therefore, the main purpose of this CSE591 Computer Visualization project is to use different visualization methods to analyze and visualize a group of datasets related to GDP. The visualization approaches we use to represent our story are Choropleth maps, histograms, line charts , bar charts and bubble plots.

I divided this portfolio into three sections. First, I explain proposed solution to the project. Next , the result of our project will be presented . Third, I will describe my contribution to our group project. Finally, I will summarize what I learned from this project.

2 PROJECT SOLUTION

2.1 Data selection

We used five fields related to GDP: GDP itself, rural and urban development, education, health and infrastructure. For each field, we collected and analyzed several data items from The World Bank Organization . Table 1 shows the data we decided to visualize in our project. After collecting the data, we used Excel as the tool to modify and filter them. These data files were stored in CSV, JSON and TSV format.

Data Field	Detail Items
GDP	GDP GDP annual growth GDP per capita
Rural and Urban Development	Urban population GDP composition
Education	School Enrollment Rate for Primary School Enrollment Rate for Secondary School Enrollment Rate for Tertiary
Health	Life Expectancy Improved Sanitation Facilities Health Expenditure Population Composition
Infrastructure	Mobile Cellular Subscription Internet Users

Table 1.1 Data selection table

2.2 Technologies

The main technologies we used are Javascript, HTML5 and CSS. D3 is a Javascript library that allows us to bind arbitrary data to a Document Object Model (DOM), and then apply data-driven transformations to the document. For example, we can use D3 to generate an HTML table from an array of numbers. Or, use the same data to create an interactive SVG bar chart with smooth transitions and interaction. D3 solves the crux of the problem: efficient manipulation of documents based on data. This avoids proprietary representation and affords extraordinary flexibility, exposing the full capabilities of web standards such as CSS3, HTML5 and SVG. With minimal overhead, D3 is extremely fast, supporting large datasets and dynamic behaviors for interaction and animation. D3's functional style allows code reuse through a diverse collection of components and plugins[3].

3 PROJECT RESULT

The result of our project is an interactive website. The first page of our system is a dashboard frame. As shown in figure 3.1, we designed and implemented the dashboard frame in this system for users to navigate to all the pages easily. Users can redirect to any page they want by just clicking the page menu [4].

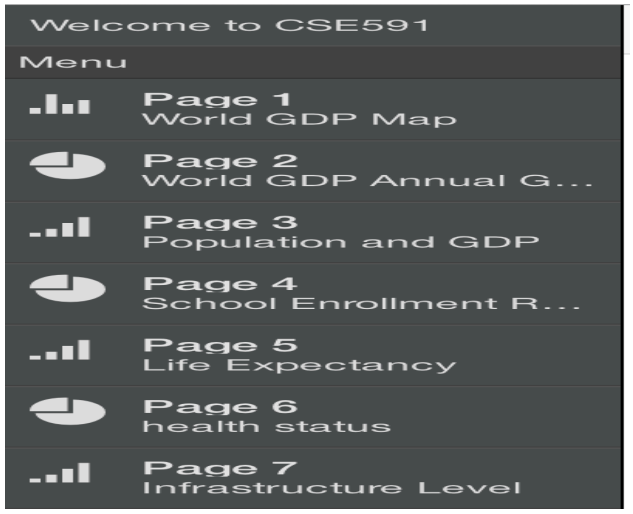

Figure 3.1 Dashboard frame

First, our project began with analyzing the GDP and GDP annual growth of worldwide countries. We provided an overview of worldwide GDP and GDP annual growth distribution by using choropleth maps.

Second, we explored the population composition of different countries. We focused on the percentage of urban and rural population of a country. In this part, we used a scatter plot to show the relationship between population composition and GDP.

Third, we analyzed the education enrollments of countries. We compared the school enrollment rate of worldwide countries. We used parallel coordinate plots to describe the relationship between education level and GDP.

Fourth, the health status of a country was also considered. The health care industry is one of the world's largest and fastest-growing industries. In many countries, the health care industry consumes nearly 10 percent of gross domestic product (GDP).

Finally, we described infrastructure distribution including Internet users (per 100 people) and telephone lines (per 100 people), since infrastructure is as important as health in the standard of living of a country.

4 Personal Contribution

In this project, I was responsible for the data analysis, visualization design and implementation of the education, health and infrastructure analysis.

4.1 Parallel coordinates Plot of Enrollment

I selected three data items from education area. These three items are school enrollment rate for primary, secondary and tertiary. I connected these three items with GDP average annual growth in this parallel coordinates plot. I divided the world into seven regions : South Asia, East Asia&Pacific, Europe&Central Asia, Middle East&North Africa, Sub-Saharan Africa, Latin America&Caribbean and North America.

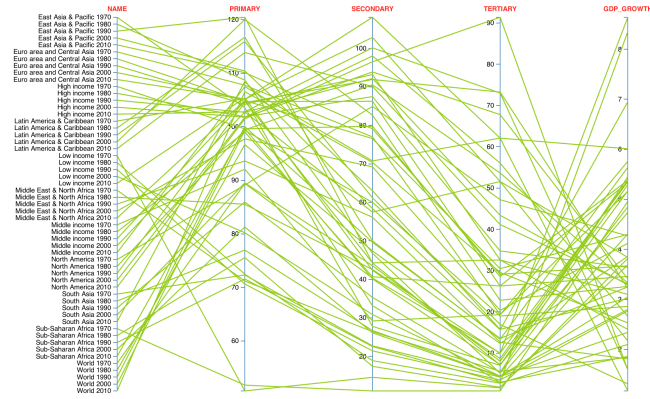

Figure 4.1 Parallel coordinates plot of enrollment

As shown in figure 4.1, I used average annual growth of the past ten years as the fifth y-axis. I set each region's item in chronological order. This allows users to make comparisons between years of the same region .

Moreover, when users move the mouse over a certain line, the line will be highlighted and detailed information of the selected line will be displayed .

4.2 Line Chart of Life Expectancy

In this part, I chose life expectancy as a factor to reflect the health status of a country. I utilized line chart to represent the life expectancy of worldwide countries.

As shown in figure 4.2, I collected life expectancy data from 1960 to 2012. I also divided 188 worldwide countries into 7 different regions: East Asia and Pacific, South Asia, Europe and Central Asia, Middle East and North Africa, Sub- Saharan Africa, Latin America and Caribbean and North America.

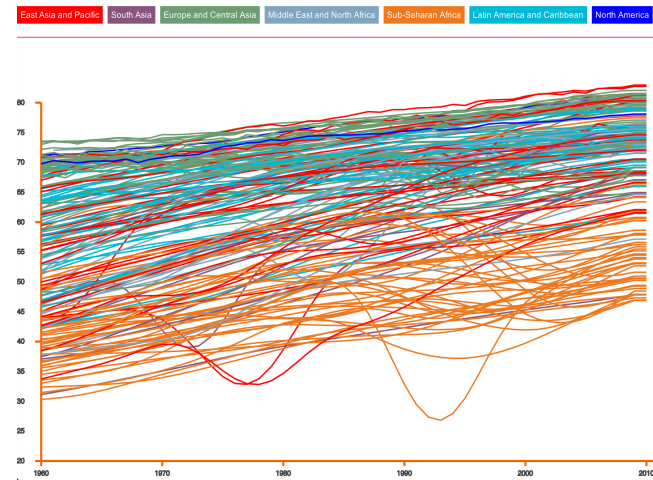

Figure 4.2 Line chart of life expectancy

In this line chart, the x-axis represents the years (1960-2012) and the y-axis represents life expectancy. Each line represents the life expectancy trend of a particular country from 1960 to 2012. A line would be highlighted when a user moves the mouse to the line and detailed information of the life expectancy of a country would be displayed. Additionally, lines would also be colored by

clicking region labels. I colored all the lines with seven different colors with respect to seven different regions. Since seven different regions are nominal data with no implied ordering, I chose a qualitative color scheme (rainbow color scale) to describe seven regions.

4.3 Histogram, Line Chart and Pie Chart of Health Status

In this part, I collected seven countries to compare their life expectancy, improved sanitation facilities (% of population with access), population composition and health expenditure (% of GDP) . These seven countries are Brazil, China, Denmark, India, Japan, USA and Zambia. Brazil, China, India and Zambia are developing countries while Denmark, Japan and USA are developed countries. Users can see a distinct and obvious difference among these seven representative countries. That is the reason why I chose these countries. Due to the huge amount of data, I only compared these countries every ten years from 1960 to 2012.

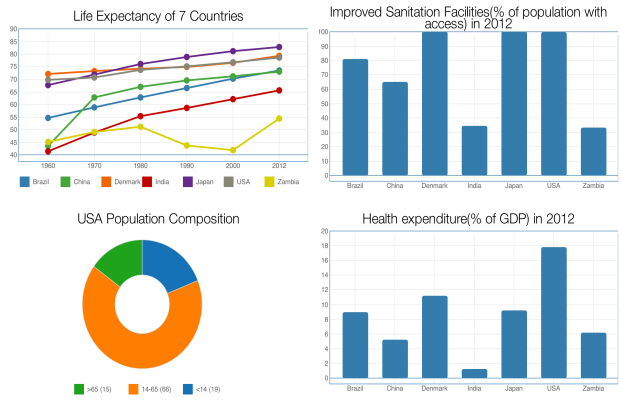

Figure 4.3 A combination of different visualizations

As shown in figure 4.3, in order to enhance the interactivity, I used a combination of different visualization ideas including histograms, line charts and pie charts to describe the data.

I used a line chart to show life expectancy, a pie chart to describe population composition and histogram to depict improved sanitation facilities and health expenditure. By clicking the dots in the line chart in a particular year, the data of the two histograms would change to that year correspondingly. Thus users can figure out the improved sanitation facilities conditions and health expenditure in that particular year. In order to get the population composition of a country in a year, users can click the histogram bin of this country and the data of the pie chart would change to that country.

4.4 Scatter Plot of Infrastructure Level

In this part, I utilized a scatter plot to present the infrastructure level of Brazil, China, Denmark, India, Japan, USA and Zambia. A scatter plot is a very useful bivariate approach to visualize discrete data values along two axes. I chose mobile cellular subscription (per 100 people) and internet users(per 100 people) as two axes. I collected the data every five years from 1960 to 2012. Each bubble in the scatter plot represents a country. Moreover, I implemented a timeline in the scatter plot to display the change of infrastructure level of these seven countries. With the change of time, the scatter plot can automatically update itself [4].

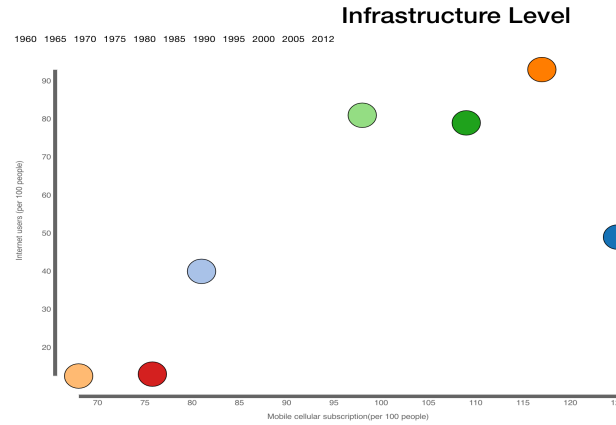

Figure 4.4 Scatter plot of infrastructure level

5 Conclusion

5.1 New knowledge learned

I have learned a lot things after finishing this project. First, I has improved my academic knowledge of computer visualization. I obtained a deeper understanding of the definition of four data types: nominal, ordinal, interval and ratio. I also learned the principles of different visualization techniques such as tree maps, histograms, fisheye views, parallel coordinates, scatter plots, linking and brushing.

Second, I acquired the fundamental skills of visualization design. I learned how to choose the proper visualization technique to present different types of data. Now, I understand that a line chart is a very efficient and effective way to present data over time and enable viewers to make predictions about the trends of data [5]. I also know that a parallel coordinates plot is a powerful for multivariate visualization [6].

Third, I improved my capability of analyzing data and programming. Now I understand how to use Excel and Matlab to analyze and filter CSV, JSON and TSV datasets. The main programming techniques I used in this project are JavaScript, D3.js , HTML and CSS. D3 is a js library that provides a lot of functions for the development of dynamic web pages. This project enables me to get more familiar with web programming development. I enhanced my ability to build visually aesthetic , effective and interactive web pages by using HTML and Javascript.

5.2 Team Memebers

Ersi Zha is a graduate student majoring Computer Science at Arizona State University. Email: ezha1@asu.edu

Jiaqi He is graduate student majoring Computer Science at Arizona State University. Email: jiaqihe@asu.edu

Yezhuo Zhu is graduate student majoring Computer Science at Arizona State University. Email: yzhu100@asu.edu

ACKNOWLEDGEMENT

The author wishes to thank Dr. Ross Maciejewski for his guidance on this project. The author also wishes to thank Jiaqi He and Yezhuo Zhu for their contributions on this project.

REFERENCES

[1] "Gross domestic product" , Wikipedia, Sep 2014. [Online]. http://en.wikipedia.org/wiki/Gross_domestic_product

[2] Nick Qi Zhu, Data Visualization with D3.js Cookbook. Packt Publishing Ltd. pp. 234-237, 2013.

[3] Mike Bostock, 2013. [Online]. Available: http://d3js.org

[4] Mullër W, Schumann H., Visualization methods for time-dependent data-an overview ,Simulation Conference, 2003. Proceedings of the 2003 Winter. IEEE, 2003.

[5] Roberton.G, Fernandez.R and Bongshin Lee, Effectiveness of Animation in Trend Visualization ,Published in IEEE Transactions on Volume:14, Issue:6, 2000

[6] Gennady Andrienko and Natalia Andrienko. Constructing Parallel Coordinated Plot for Problem Solving. German National Research Center for Information Technology , 2001