A Story Behind GDP

Jiaqi He, Ersi Zha, and Yezhuo Zhu, Student, *ASU*

Abstract—GDP is widely mentioned in our daily life, such as TV, newspapers, and magazines. However, it is more than a set of values. There are a large amount of information are hidden behind these numbers. This paper introduces a system, which focuses on telling a story behind GDP via several frames generated by D3.js. These frames visualized 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. Detail description is presented illustrating the design ideas, the features and the implementation process results of this system. Moreover, this paper outlines the instructions on how to use this system and the information found by using this system.

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

1 INTRODUCTION

AS economic globalization rapidly develops, Gross Domestic Products (GDP) has become one of hottest topics in recent years. But sometimes GDP cannot comprehensively and accurately reflect the living standard of a country because the living standard of country depends on many aspects such as rural and urban development, health status, poverty situation, infrastructure and so on. Today many people are still confused about the relationship between GDP and these aspects. The main purpose of our project is to explore how GDP impacts the living standard of a country and make a comparison of some major countries in the world. We plan to use various visualization approaches such as Choropleth map, histograms and bubble plot to represent our story.

In this paper, we will mainly explain the related work, data selection and visualization design ideas. After watching Hans Rosling's TED lecture and studying the economic knowledge, we decided to select five aspects to make connection with GDP: Urban Development, Education, Health and Infrastructure. In conclusion, we will discuss the breakdown and eleven reference papers.

2 RELATED WORK

In recent years, many works have been done in this field. Swedish medical doctor, academic and statistician Hans Rosling used software Trendalyzer to change tedious time series of development data into fascinating moving graphics [1]. Bubble-shaped country icons were applied in his World Health Chart. The size and the color represented the population and the continent respectively. The user can enjoy an animated change to a new position when he changes settings. Also, he applied Dollar Street graph accompanied with some video documentation to illustrate daily income per person in the world. An interactive graph was used to show the world income distribution and he selected several countries' statistics between 1970-1988 to display.

In our system, we refer the idea of bubble chart and animation in Rosling's research work. However, we added many other ways to tell our story

3 DESCRIPTION OF SYSTEM

3.1 Data Selection

We tried to tell the story behind GDP in these five aspects, GDP itself, rural and urban development, education, health and infrastructure. For each field, we fetched serveral data items from World Bank Organization (as shown in Table 1).

Table 1

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 3.1 shows the data we are going to visualize in our project. Since the purpose of our project is to explore and compare the living standard of different countries in the world, for this project, we will focus on seven major typical countries in the world to make comparisons. It is obvious that the quality of life in a country is closely relevant to its economic development level and GDP is one of the most important factors to measure the economic development of a country. So our story will begin with analyzing the GDP and GDP annual growth of worldwide countries. We will give people a comprehensive overview and summary of worldwide GDP.

However, GDP cannot comprehensively evaluate the development of a country. For example, high GDP does not mean high-level living standard. Also, there are a lot of factors that can influence GDP. Therefore, we will use our selected data to explore the story behind GDP.

Firstly, we will examine the population composition of different countries, which is assumed to have a close relationship with GDP. It will be very interesting to analyze how different population compositions contribute to different GDP. We will figure out the percentage of rural population and urban population of seven major countries. Then we will explore the population proportion of different age range of those major countries. The proportion of population ages 15-64 of a country is a very important factor to decide the country's GDP level for the reason that population aged 15-64 would definitely make more contributions to GDP than other age groups population.

Secondly, we will examine the education situation of above countries. We want to compare the school enrollment rate of worldwide countries. We want explore the relationship between education level and GDP.

Thirdly, health status of a country will also be considered. The health care industry is one of the world's largest and fastest-growing industries. In many countries, health care industry consumes nearly 10 percent of gross domestic product (GDP). In this part, we are going to make a comparison of health status including life expectancy of seven major countries.

Last, we will give statistics of infrastructure distribution, such as Internet users (per 100 people) and telephone lines (per 100 people), since infrastructure is as important as health to decide the living standard of a country.

4 System Desgin and Implementation

In this section, we will discuss the design and implementation process of our system. We broke this part into different sections with respects to different visualization ideas. The software we used to code and implement our project is Webstorm and the primary technologies we used to construct our system are JavaScript, HTML, CSS and D3. We used CSV,JSON and TSV data types to store our system data. Next, we will explain the details of system design and implementation with respect to different visualization design ideas. For D3.js part, we referenced some open-source libraries in the Internet and the reference book [2][3].

4.1 Choropleth of GDP

In this choropleth, we visualized the Global GDP for ten years. Since GDP do not change drastically from one year to the next year, we did not pick up data every year. In order to see GDP change in a more obvious way, we pick up data ever five years from 1965 to 2010 to generate this Choropleth. The following figure 4.1 shows our choropleth map of GDP.

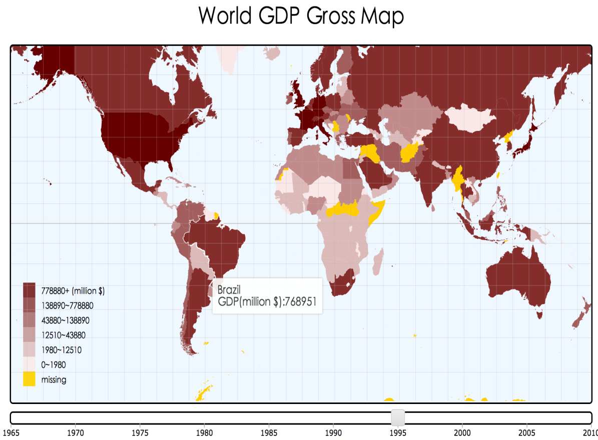

Figure 4.1 World GDP Gross Map

As for the color scheme, we chose sequential color scheme (as shown in Figure 1), which is a good choice for interval data type. Based on what we learned in geographic visualization lecture, the typical cartographic rule of thumb is 5-7 classes. So we chose 6 as the number of bins in our map. Darker color means larger GDP and lighter means smaller GDP. Also, some countries do not have data for several years, for example, Russian Federation do not have data before 1990 for it was part of Soviet during that period. For these situations, we set yellow to represent these missing data. Then we chose quantile classification in this visualization for the reason that it is easy to compute and each class has the same percentage of observations. In order to enable users to make comparisons between years, we utilized a uniform color legend for all the ten years' choropleths. Additionally, we provided users with several ways to interact with this choropleth, such as the timeline slider, dynamic tooltip, zoom in and zoom out and drag and move.

As for the implementation process, we used Matlab, Python, HTML, CSS and D3.js to accomplish this visualization. After fetching GDP data from Word Bank, we selected the ten years data we needed. Then we imported these ten years' data into Matlab and used it to generate the quantile values for the color legend. Next, we wrote a piece of Python code to match the country names in GDP data with those in a CSV format file, which stores the data about countries' boundary information. In the meanwhile, the Python code also assigned each country an appropriate color code based on its GDP value and the quantile values we got in the last step. After that, we got several JSON files containing each country's information. Then we imported these data files in our D3.js code combined with HTML and CSS to generate this GDP choropleth.

4.2 Choropleth of GDP Annual Growth

Since the choropleth of GDP cannot expose much information behind GDP for the reason that in most situations, larger area or larger population means larger GDP. This makes GDP less meaningful than GDP annual growth when we try to explore how the world develops with time going on. Therefore we created another choropleth for GDP annual growth. In order to expose information as much as possible, we picked up every year's data from 1964 to 2010 to generate this annual growth choropleth. Also, users can use the timeline slider to decide the year to display. The following figure 4.2 shows the world GDP growth Map.

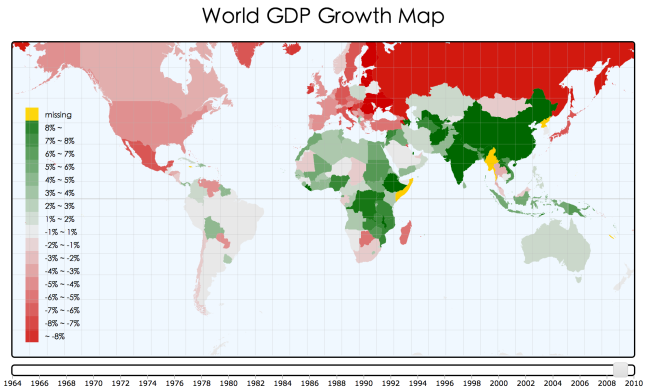

Figure 4.2 World GDP Growth Map

Compare with the first GDP choropleth, we almost applied the same template. However, there still exist some differences. Apart from we imported more years' data than the first choropleth, we also chose diverging color scheme instead of sequential color scheme on the basis of the fact that the annual growth could be positive or negative. Since the opponent process theory of process suggests that black-white, yellow-blue and red-green are the best three choices for human perception [4], we chose red/green scale for our color legend. Green represents positive annual growth and red represents negative annual growth. Additionally, we chose equal interval classification for this color legend. Unlike GDP, most countries' annual growths do not vary a lot, they are usually between -8%~+8% instead. So we set each percent as a single class in this range. Particularly, the darkest green represents annual growth more than 8% and the darkest red represents annual growth less than -8%.

As for the implementation process, we used almost the same set of tools to generate the second choropleth. The only difference is that we did not use Matlab during the implementation process, because this choropleth needn't quantile values.

4.3 Scatter Plot and Histogram of Urban Development

As shown in figure 4.3,in this visualization, we tried to visualization the relationship among GDP per capita, the urban population ratio and GDP with the scatter plot. We wanted to transform these data of three dimensions into the geometric to help the researcher observe the relationship among them[5]. After discussion, we decided to use scatter plot. The X-axis represent GDP per capita, Y-axis represent the urban population ratio, every circle represent a country and the area of each circle represent GDP of this country. The histogram below shows the ratio of the different industries in GDP such as the trade, the agriculture, the manufacturing, the industry and the service. The X-axis represent the ratio of the industries in GDP and the Y-axis represent the name of the different countries. When the circle is clicked, the ratio of the different industries in GDP of this country will be shown in this histogram.

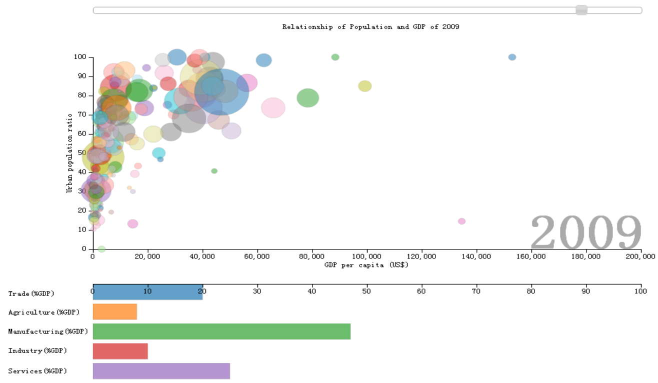

Figure 4.3 Scatter plot&&histogram of Urban

The original data are very large. They cover hundreds of countries in more than 20 areas of the urban development and GDP. But we focus on the urban development ratio and GDP per capita. By doing this, the pattern will be more apparent and easier for the analysts to find the interesting subsets of data[46]. We collect data of ninety countries from 1961 to 2010 and this data will be dynamically represented[7] by the scatter plot year by year. By controlling the silder, the data of the specific year can also be checked. In this way, the data will be represented statically[7].

For the implementation of this visualization, we used the D3.js library, which provides a lot of functions for development of the dynamic web pages. For the multi-variables plot, the number of the countries seems to be too large, so we hyalinized the circle transparent to make the viewers see all the circles. The label at bottom right corner shows the current year and the name of the specified country represented by the circle clicked. The data was stored in the file of the csv format and D3.js was used to import this data.

4.4 Parallel Ordinates Plot of Enrollment

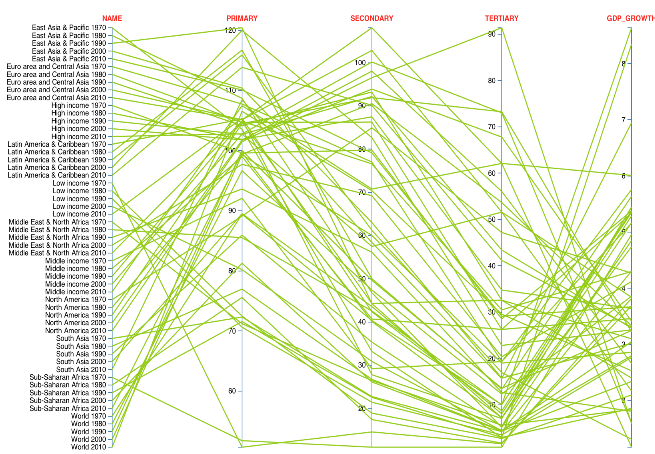

Figure 4.4 Parallel Ordinates Plot of Enrollement

As shown in figure4.4, in this visualization, we tried to tell the story in the education area. We picked up data items of school enrollment rate for primary, school enrollment rate for secondary, school enrollment rate for tertiary and GDP average annual growth of the past ten years for seven different regions in the world (South Asia, East Asia&Pacific, Europe&Central Asia, Middle East&North Africa, Sub-Saharan Africa, Latin America&Caribbean and North America) of different years (1970, 1980, 1990, 2000 and 2010). Since parallel coordinates plot is a powerful tool for multivariate visualization, we chose it to visualize the data in education field[8].

There are five coordinates in this visualization. Since the order of the parallel coordinates plot has great impacts on the resultant visualization, we thought carefully before we set the order these five coordinates. At last, we decided to layout the parallel coordinates as shown in Figure. The first coordinate is mapped to different regions with different years. The second to the fourth coordinates are mapped to school enrollment rate for primary, secondary and tertiary respectively, which reflects the order of different educational stages. At last, we put average annual growth of the past ten years on the fifth coordinates. Particularly, we set each region items in chronological time order in order to enable users to make comparisons between years of the same region. Also, when the user move the mouse over a certain line, the line will be highlighted and all the rest lines will become grey. In the meanwhile, a tool tip will pop up to display the information in these five coordinates for the selected line.

In the implementation process, we first used Excel to selected data items we needed and calculate the average annual growths of each ten years from 1960 to 2010 (1961~1970, 1971~1980, 1981~1990, 1991~2000 and 2001~2010). Then we input these selected data items into TSV file manually, which will be used in our D3.js code. At last, we implemented this parallel coordinates plot with HTML, CSS and D3.js.

4.5 Line Chart of Life Expectancy

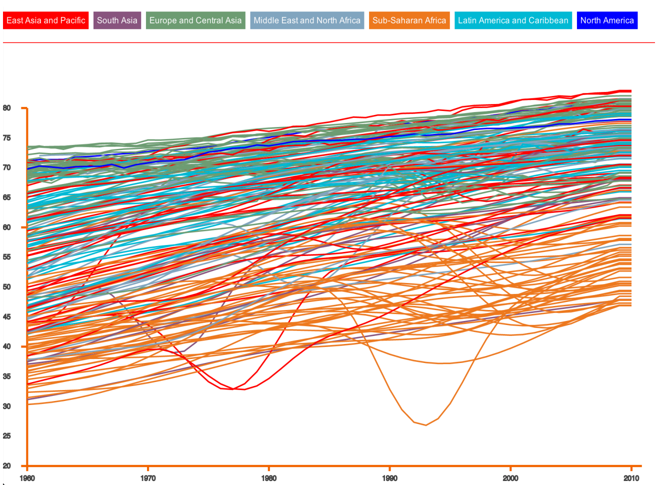

Figure 4.5 Line chart of life expectancy

In this part, we utilized line chart to describe the life expectancy of worldwide countries. The reason why we chose line chart to display data is that line chart is a very efficient and effective visualization tool to show data over time and enable viewers to make prediction about the trends of data[9]. Since life expectancy is a crucial factor to reflect and judge the health level of a country, we decided to depict the data of life expectancy.

We collected life expectancy data from 1960 to 2012 and 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. There are two CSV files related this line chart. First, we used a country-regions.csv file to associate country code with region code. So we can figure out which region a country belongs to. Then we used lifeExpectancy.csv file to store the life expectancy of worldwide countries

In our line chart, X-axis represents years(1960-2012) and Y-axis represents life expectancy. Each line represent the life expectancy trend of a particular country from 1960 to 2012. In order to provide interaction between viewers and data, when we move mouse to a line in our line chart, this line would be highlighted and detail life expectancy information of this country would be displayed. Moreover, the lines in our line chart would be colored by clicking region labels. We colored all the lines with seven different colors with respect to seven different regions, therefore, we can make comparisons among these seven regions and read more useful information from this line chart. Since seven different regions are nominal data with no implied ordering, we chose qualitative color scheme (rainbow color scale) to describe seven regions.

Consider the implementation process, we first normalized the data from worldbank organization and imported these data to our project in Webstorm. We used D3.js as the tool to draw line chart in HTML context and we used CSS to style the layout, colors and font of our graph

4.6 Histogram, Line Chart and Pie Chart of Health Status

After the overview of life expectancy conditions of all worldwide countries, we selected seven typical countries to further compare their health status according to four fields: 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 four developing countries while Denmark, Japan and USA are three developed countries. So we can see a distinct difference among these seven countries. That is the reason why we chose these seven countries. Due to the huge amount of data, we only compared these countries every ten years from 1960 to 2012.

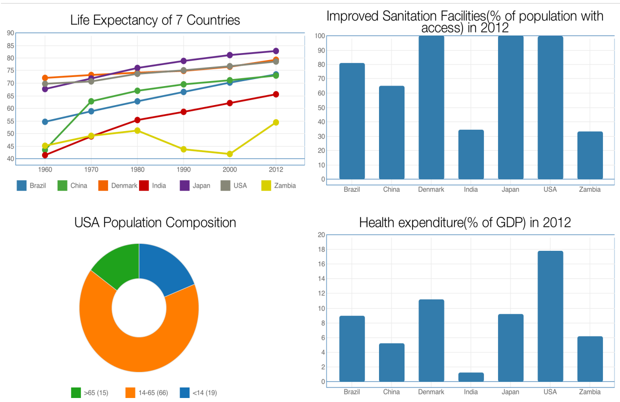

Figure 4.6 A combination of different visualizations

As shown in figure 4.6, in order to enhance our visualization, we decided to use a combination of different visualization ideas like histogram, line chart and pie chart to describe our data. We connected these graphs with each other to make system more interactive[10].This would definitely offer viewers a more clear and intuitional description of data.

We used line chart to show life expectancy, pie chart to show population composition and histograms to show improved sanitation facilities and health expenditure. By clicking the dots in line chart in a particular year, the data of two histograms would change to that year dynamically. So we can see the improve sanitation facilities conditions and health expenditure in that particular year. In order to get the population composition of a country in a year, we can click the histogram bin of this country and the data of pie chart would change to that country correspondingly. As we did before, we still used csv file to store data and used D3.js ,CSS and HTML technologies to implement visualization.

4.7 Scatter Plot of Infrastructure Level

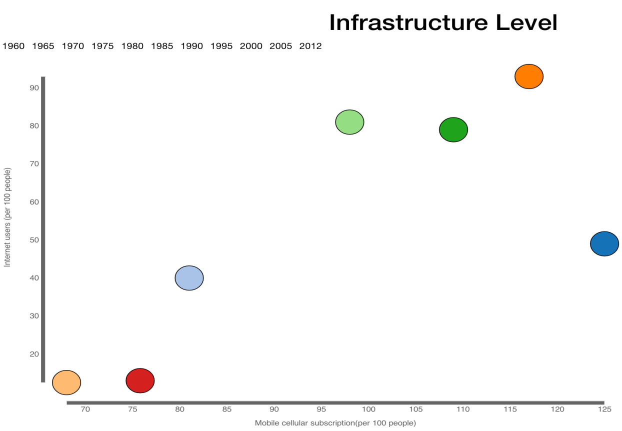

Figure 4.7 Scatter plot of infrastructure level

In this part, we utilized scatter plot to show the infrastructure level of the same seven typical countries in last section. Scatter plot is a very useful bivariate approach to visualize discrete data values along two axes. We selected mobile cellular subscription(per 100 people) and internet users(per 100 people) as two axes because these two are very important attributes to judge the infrastructure level of a country.We collected the data every five years from 1960 to 2012.

In this scatter plot, X-axis represents Mobile cellular subscription(per 100 people) and Y-axis represents Internet users(per 100 people).Each bubble in the scatter plot represents a country. Moreover, in order to display the change of infrastructure level of these seven countries, we add a timeline in our scatter plot. So with the change of time, our scatter plot can automatically update itself. As we did before, we still used csv file to store data and used D3.js ,CSS and HTML technologies to implement visualization.

5 Results

Our system is very easy to use. As shown in figure 5.1, we designed and implemented a dashboard frame in this system for users to navigate all the pages easily. System would redirect to any page you want by clicking the page inventory[11].

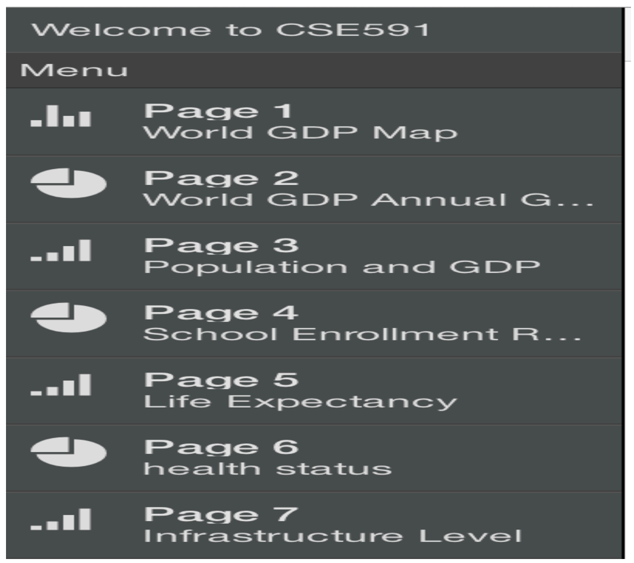

Figure 5.1:Dashboard frame of our system

For the first page, the world GDP map, timeline slider below the choropleth enables users to change the year to display. When users move the mouse over a country, a tooltip will appear to show the country name and its GDP value. Also, users can use the scroll wheel to zoom in and zoom out. By clicking the mouse, users are able to drag the whole view to positon he wants. When change from year to year, we can easily find that, most contries's colors become darker and darker, which means their GDP become larger and larger.

For the second page, world GDP annual growth map, the interactive ways implemented in this map are the same as the first choropleth. However, we can find more useful information from that. Dragging the slider to different positions, we could find different stories behind GDP. For example, when we set the position to 1998, most countries' in East Aisa are red, which was caused by Asian Finacial Crisis last century. Another example, when setting the position to 2009, the Global Financial Crisis that year caused many countries all over the world becoming red. In the next year, 2010, most contries recovered from the economy recession. However, Greek government experienced a debt crisis, which caused its negative growth in 2010.

For the third page, the scatter plot shows the relationship between GDP per capita and urban population. Users can see the automatic changes of the scatter plot from 1961 to 2010 by pressing start/stop button. Users can also drag the slider to see the data of the countries at the specific year. Circles in the scatter plot represents different countries.When users move mouse to a circle, the name of the country will be shown at the bottom right corner. By clicking the circle, the histogram under the scatter plot will display the accurate percentage of different fields in GDP including trade, agriculture, manufacturing,industry and services.

For the fourth page, the parallel coordinates plot provides users lots of information of schoole enrollment rate with respect different regions in different years. When users move mouse to a line ,the line will be highlighted and detail information of school enrollment rate of that region in a particular year will be displayed. As we divided worldwide countries into seven different regions and high,middle and low income countries, we can not only compare the school enrollment rate of these seven regions but of those countries with different income level.

For the fifth page, users can explore the change of life expectancy of worldwide countries from 1960 to 2012. At first, uses can easily figure out the life expectancy trend of all the countries by analyzing lines in the chart. In order to get detail information of a country ,users can move mouse to the line they are interested in. Furthermore, we offered users the function to compare the life expectancy of different regions. We also divides worldwide countries into seven regions. Users can make a comparison of different region by clicking region labels to color the lines with different colors according to different regions.

For the sixth page, users can compare the health status of seven countries includes Brazil, China,Denmark, India, Japan, USA and Zambia. Users can explore the life expectancy, improve sanitation facilities, population composition and health expenditure in a combination of different visualization graphs. In order to provide a more interactive environment between users any system, we made these graphs connected to each other.

For the last page seven, we made a bubble plot to show the infrastructure level of seven major countries: Brazil,China, Denmark,India,Japan,USA and Zambia. We offered a timelines label in this system. By clicking the label of different years, the bubble plot will dynamically change to that year. Country's name will be displayed,if users move mouse to the bubbles.

6 Conclusion and Future Work

Due the limitation of time, we cannot make every perfect.There is plenty of room for improvement in our system. In the further, we want to add more visualization ideas and provide more interactive environment between users and system.

In the two choropleth maps, more interactive ways could be added. For example, when users move the mouse over a country, a histogram could appear to show how GDP or GDP annual growth change over years and highlight the point of that year[12]. In parallel coordinates plot, a set of buttons could be added to enable users to choose which years' line to display. This can make it easier to compare education rate between regions in the same year. Also, we can replace poly-lines with smooth curves to enable individual data elements to be traced[13].

For the scatter plot and the histogram shown in the graphic 4.3, it can help the viewer analyze the impact of the urban population on GDP per capita. In addition, by exploring the ratio of the different industries in GDP, we can also infer the reason why the high urban population has a positive impact on the GDP per capita. Although we did our best to make this visualization more perfect, there are still some problems with this visualization. Firstly, since the number of the countries is too large, we can not find a proper method to assign the color to these countries. Secondly, the industry data of 40 years from 90 countries are too large and a lot of data are missing. Collecting data is too hard for one or two person to complete. So we just use the industry data of 2009 for analyzing. In future, we would continue to improve this visualization to make it more robust and make data more intact.

References

[1] Rosling H, Rosling R A, Rosling O. New software brings statistics beyond the eye[J]. Statistics, Knowledge and Policy: Key Indicators to Inform Decision Making. Paris, France: OECD Publishing, 2005: 522-530

[2] Murray S. Interactive data visualization for the Web[M]. " O'Reilly Media, Inc.", 2013.

[3] Zhu N Q. Data Visualization with D3. js Cookbook[M]. Packt Publishing Ltd, 2013.

[4] Hoeber, O., Wilson, G., Harding, S., Enguehard, R., & Devillers, R. (2011, March). Exploring geo-temporal differences using gtdiff. In Pacific Visualization Symposium (PacificVis), 2011 IEEE (pp. 139-146). IEEE.

[5] Chen M, Ebert D, Hagen H, et al. Data, information, and knowledge in visualization[J]. Computer Graphics and Applications, IEEE, 2009, 29(1): 12-19.

[6] Derthick M, Kolojejchick J, Roth S F. An Interactive Visualization Environment for Data Exploration[C]//KDD. 1997: 2-9.

[7] Mullër W, Schumann H. Visualization methods for time-dependent data-an overview[C]//Simulation Conference, 2003. Proceedings of the 2003 Winter. IEEE, 2003, 1: 737-745.

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

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

[10] John V.Carlis and Joseph A.Konstan, Interactive visualization of serial periodic data, Published in Proceddings of the 11th annual ACM symposium on User interface software and technology ,Pages 29-38 ,1998

[11] Marc A.Smith, Andrew T.Fiore Visualization components for persistent conversations. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.Pages 136-143,2001

[12] Diansheng Guo, Jin Chen and MacEachren. A, A Visualization System for Space-Time and Multivariate Pattern, Published in Visualization and Computer Graphics, IEEE Transactions on Volume:12,Issue:6, Nov.-Dec.2006

[13] Graham, Martin, and Jessie Kennedy. "Using curves to enhance parallel coordinate visualisations." Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on. IEEE, 2003.