

Project Proposal for Information Visualization

Section 1: Basic Information

Project Title: Trend of population and GDP in each province of mainland China

Team members & Net IDs: Jia Zeng (jz4667) Qingyang Li (ql2048) Ruixiong Yan (ry2129)

Section 2: Overview

China has undergone a lot of changes in recent several decades, especially after the reform and opening-up in the 80s. The Chinese market is exposed to eyes all over the world. Most clever investors will pay attention to China and hope to learn more. Population and GDP are two main indexes to reflect the society and economy in a district. We plan to propose a visualization system to show the population migration and economic growth in provinces and reveal the story behind these sharp changes in China. The viewers can quickly obtain the key information, relation between population and GDP, from our visual work.

Section 3: Description of the dataset and processing

We are planning to use the following datasets:

(1) Population of China (2020 and historical)

URL: <https://www.kaggle.com/datasets/anandhuh/population-data-china>

We will visualize the following variables: 1) Year: years from 2020-1955; 2) Population: population in the respective year; 3) Density: population density (population per square km).

(2) China's GDP in Province

URL: <https://www.kaggle.com/datasets/concyclics/chinas-gdp-in-province>

This dataset contains the GDP of each province in mainland China between 1992 and 2020.

We will visualize all the data and let users select the year/province for detailed observation.

(3) Population of each province of China

URL: <https://www.renrendoc.com/paper/126295415.html>

This dataset contains the population of each province in mainland China between 1949 and 2018. Data for the years 2019 & 2020 will be obtained from the National Bureau of Statistics (NBS) of China. All the attributes will be visualized.

(4) Map of China

URL: <https://github.com/deldersveld/topojson/>

We will use the map of China (in TopoJSON format) to implement geographical views.

We are planning to process the dataset in the following ways:

We will select the data of the year 1992-2020 from Dataset (1-3). We will also manually type in the data of 2019-2020, after obtaining it from NBS.

Dataset (3) is in xls format, and we plan to convert it to csv.

We will derive the attribute of "GDP per capita" from joining dataset (2) and (3). We will also derive the attribute "proportion of population in China of province" directly from dataset (1).

Section 4: Usage scenarios & tasks

Chris is an investment manager of a global company, which expects to expand their branches into China. His job is to find the best place to build a new factory for the China branch, since he is supposed to fully know the basic situation and development trends of China. Our project can draw a picture for him. Firstly, Chris will see the geo-maps at the top of the page. The saturations for each province can represent the GDP and population data in a unified scale. So Chris will easily compare the differences between all provinces to help him make decisions. Additionally, he can use the range slider to choose years between 1992 to 2020, the change of color in the map provides straightforward and dynamic results on the migration direction and economic growth rate over these years.

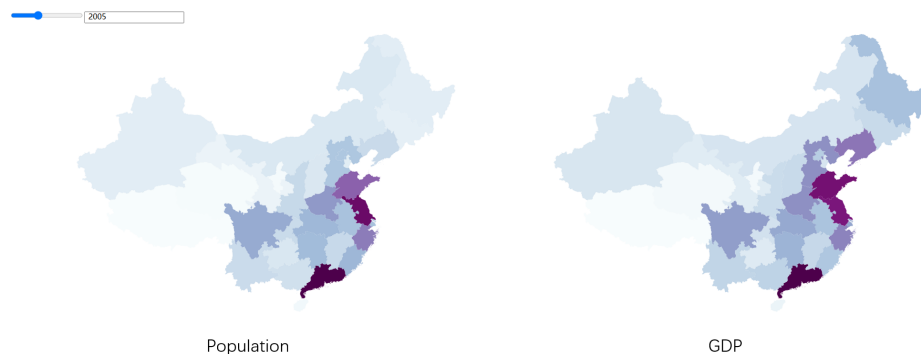
Secondly, he can see bar charts including the related data of all provinces. The bar chart with numbers will be more accurate, he can make use of it to compare the subtle differences between any two places.

The above contents are more relevant to space comparison. The final part will turn to focus on the time. If Chris is interested in the background and history of a specific place, he can click the dropdown to select that province, then he can see the line chart displaying the complete development over the years. By our charts that contain rich information, Chris can discover what he needs and optimize the strategy for his company.

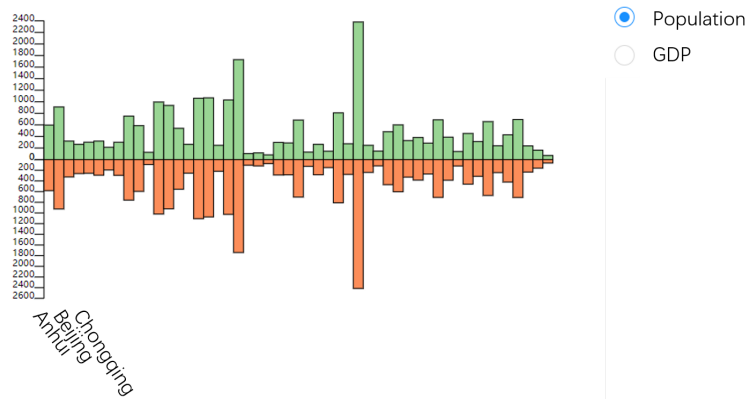
Section 5: Description of your visualization & sketch

We plan to draw three layouts, including a geo-map, a bar chart and a line chart in one single-screen dashboard. All views belonging to the same layout are linked highlighting.

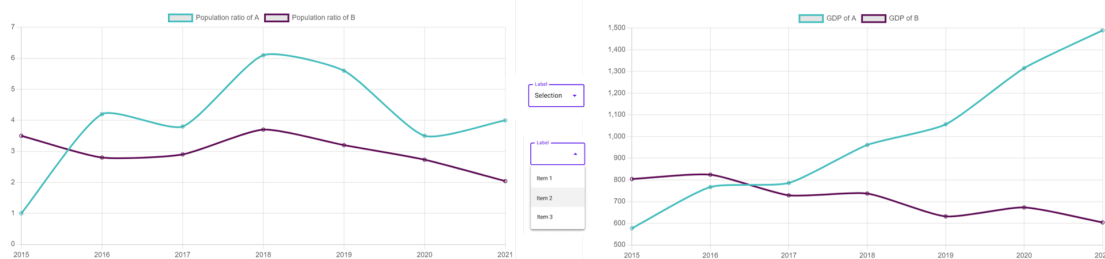
On the top of the web page, we plan to draw two geo-maps and a range slider UI widget. Users can use the range slider to select the year they are interested in and view two geo-maps. The range slider also controls the selection of the year of the bar chart in the next paragraph. One geo-map shows the proportion of the resident population of different provinces in the total population of the country in the year selected by the user. The other one shows the total GDP of different provinces in the same year. The corresponding province will be highlighted on both two maps when users hover over a province. There will also be a tooltip displayed next to the mouse. The tooltip contains information including the name of the province, the resident population of the province, the proportion of the resident population of the province to the total national population, and the total GDP of the province. It is worth mentioning that this part is our "innovation view component". We visualize two different attributes on two identical maps, which can better enable users to find out the potential relationship between the two attributes by comparing the two geo-maps.



After that, we will draw a bar chart with an upper part and a lower part under the geo-maps. The upper bar chart shows the total resident population of each province and the lower one shows the per capita GDP of each province. The upper bar chart and the lower bar chart are linked highlighting. Beside the bar chart is a radio button. All bars in the bar chart will be sorted by the total resident population of each province by default. However, users can choose whether to sort according to the population of each province or the per capita GDP of each province through the radio button next to the bar chart.



On the bottom of the web page, there will be two line charts with two dropdowns. The line chart on the left shows the trend of the proportion of the resident population of the selected provinces in the total population of the country over time. The line chart on the right shows the trend of the per capita GDP of the selected provinces over time. There will be two dropdowns between the two line charts. The reason for designing two dropdowns is that users can select two provinces they are interested in for comparison.



Section 6: Work breakdown and schedule

Milestone 1: Apr 21 – Apr 27

In this phase, each team member will be designated to implement one of the three views described above separately. The breakdown is as follows:

Li, Qingyang: geo-map, range slider and tool tip

Yan, Ruixiong: line chart & drop-down menu

Zeng, Jia: bar chart & radio button

Milestone 2: Apr 28 – May 04

In this phase, the team as a whole will combine these views to form an integrated visualization solution. Integration tests will also be carried out.

Milestone 3: May 05 – May 11

In this phase, the team as a whole will examine the visualization and write the final report. Each team member will describe their own implementation in the report. Besides, the PPT slides and demo will also be prepared for the final presentation.