

CS6630 Visualization Fall 2017 Project
A Mirror of History

Yanqing Peng, Yuwei Wang

November 10, 2017

Get started! (Nov 4)

After spending a lot of time on data collecting, we successfully collected the raw data of the historical country borders. The data we collected come from two different data sets:

- For maps after World War II (1946-2017), we collect from www.thenmap.net. Although this website doesn't provide the off-the-shelf dataset, one of its APIs supports querying for the world map at a specific year after WW2. The API we used looks like following:

```
wget http://api.thenmap.net/v1/world-2/geo|data/1946?data_props=name|wikida
```

This query retrieves the world map at 1946 with the metadata of country name as well as the wikidata id for the country. The wikidata ID is essential for future data collection such as population.

- For maps from 2000BC to 1945, we find the dataset in a cartography forum (<http://www.cartotalk.com>). The author is anonymous. The dataset only contains a few years. That's reasonable enough – we never expect to find a dataset contains all data from 2000BC to 2017AD!

For other data (e.g. population), it seems promising to retrieve them from Wikidata. We plan to explore the possibility for that. The first dataset already includes the Wikidata ID for all countries. We may have to do some data processing on the second dataset to add the Wikidata ID information to it.

Implementing the map (Nov 5)

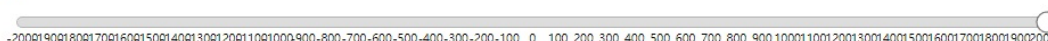
We implemented a prototype of the map chart and the year chart. It looks like the following (Figure 1):

Project

Name: Yanqing Peng; E-Mail: yq.peng@utah.edu; UID: u1076076

Name: Yuwei Wang; E-Mail: u1140944@utah.edu; UID: u1140944

2017



World Map



Figure 1: Current state of our project, Nov 5.

For the year chart, there is a slide bar and a circle which we can drag to choose the desired year. When the choice of year is updated, the map chart read the dataset with the nearest year before the selected year, and draw it on the screen. The projection we currently use is `d3.geoPatterson()`.

Adding label is quite challenging. If we simply put all labels on the map, then there will be a mess in Europe. Currently we only show the labels for countries with areas larger than a threshold. That doesn't work for countries with large area but extremely long names (e.g. Democratic Republic of the Congo). Whether or not a label is shown should depend on both country area and the length of country name. We should think carefully about it.

The location of the labels is another issue. Currently we put the label at the centroid of the corresponding path. But it doesn't work for some countries like the U.S. The label of U.S. is actually located inside Canada because of Alaska. We have to find a smarter way to locate the labels.

Layout design, and more on label placement (Nov 6)

We finished a prototype of the info panel. We also improved the label placement. As shown in Figure 2, now it looks better, isn't it?



Figure 2: Current state of our project, Nov 6.

As shown in the above figure, the information of the capital of a given country has already been successfully shown. We use Wikidata to retrieve all such information. The query uses SPARQL. As an example, if we want to know the current capital of Russia, the language looks like the following:

```
SELECT ?name WHERE {  
    wd:Q159 wdt:P36 ?link.  
    ?link rdfs:label ?name.  
    filter (lang(?name) = "en").  
}
```

Where wd:Q159 is the wikidata id of Russia, wdt:P36 is the wikidata id of the property “capital is”, and the variable ?link becomes the capital of Russia. Note that ?link itself is also a wikidata id, which is Q649 in this case. We want to know what Q649 is, so we use the second line in the where clause to retrieve the its label, and use the third line to restrict the answer to be in English.

For the label placement, we devised a smarter way to put the labels. We first decompose each country into continuous polygons, and pick the polygon with largest area (e.g. for the U.S., only the continent part will remain). Then we calculated the longest segment within the polygon, and place the label on that line. The font size will be dynamically adjusted according to the ratio between the segment length and the label length.

Information retrieval ready (Nov 7)

We finished most information retrieval part, as well as the css style for selected country. We also did a little improvement on label placing: instead of just placing the text on the diameter of the polygon, now the text is placed on a curve going from one end of the diameter to the centroid, then to the other end of the diameter. As shown in Figure 3, it significantly improve the effect for countries with trapezoidal shapes such as U.S. and China, compared to Figure 2.

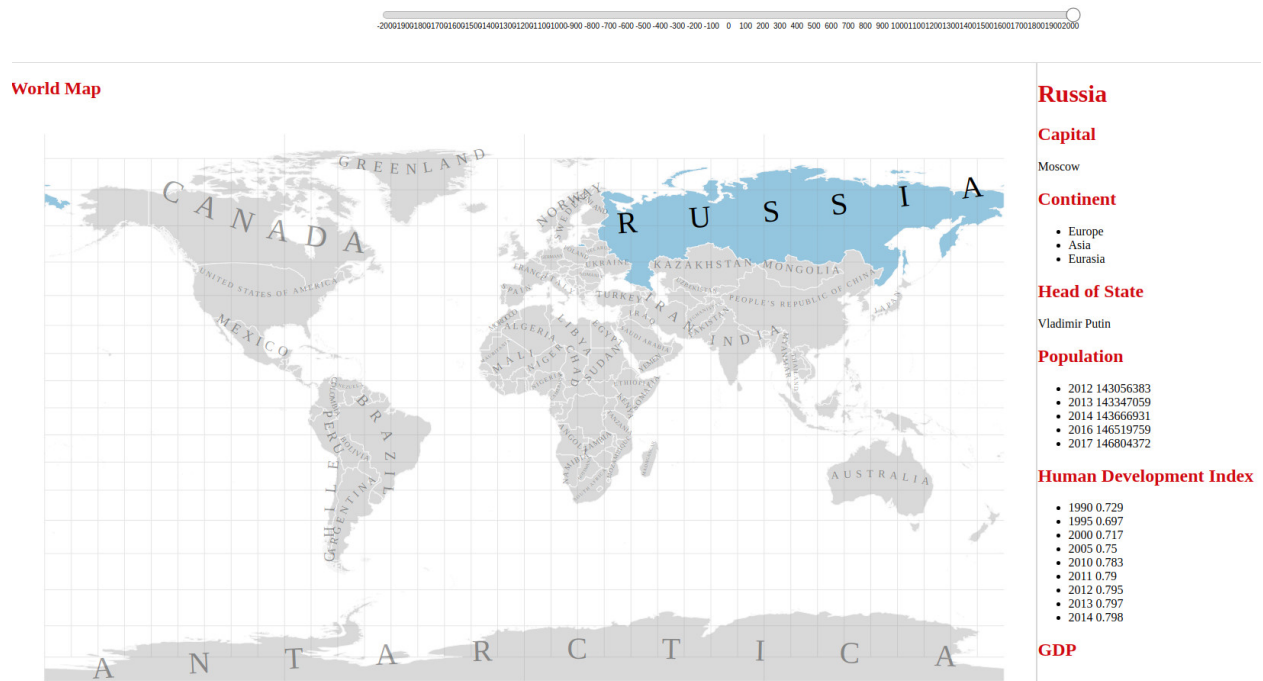


Figure 3: Current state of our project, Nov 7.

InfoPanel design (Nov 8)

For info panel, currently there are eight parts including selected country name, capital, continents the country belongs to, head of state, population, human development index, territory and description from Wiki. We retrieved all the data from Wikidata and just put them as original way without any virtualization. In this way, we found that some data are redundant so deduplicating is necessary. For example, the continents Russia belongs to include Europe, Asia and Eurasia but Eurasia means Europe and Asia. In addition, we decided to label each tile when the mouse hovering over it since the population data are quite sparse in some areas.

For the description from Wiki, we first tried to bind the source of link with the direct search result from Wikidata. However, in this way irrelevant information takes too much space like the nav bar of Wikipedia. We took into account the limitation of the page size and got rid of these information by adding "printable=yes" at the end of the link.

Visualizing the InfoPanel design (Nov 10)

We were busy these days. The only progress we made in these days is that we tried to visualize the data in the information panel. As shown in Figure 4, now the population is shown as a line chart, and the HDI is shown as a tile chart.

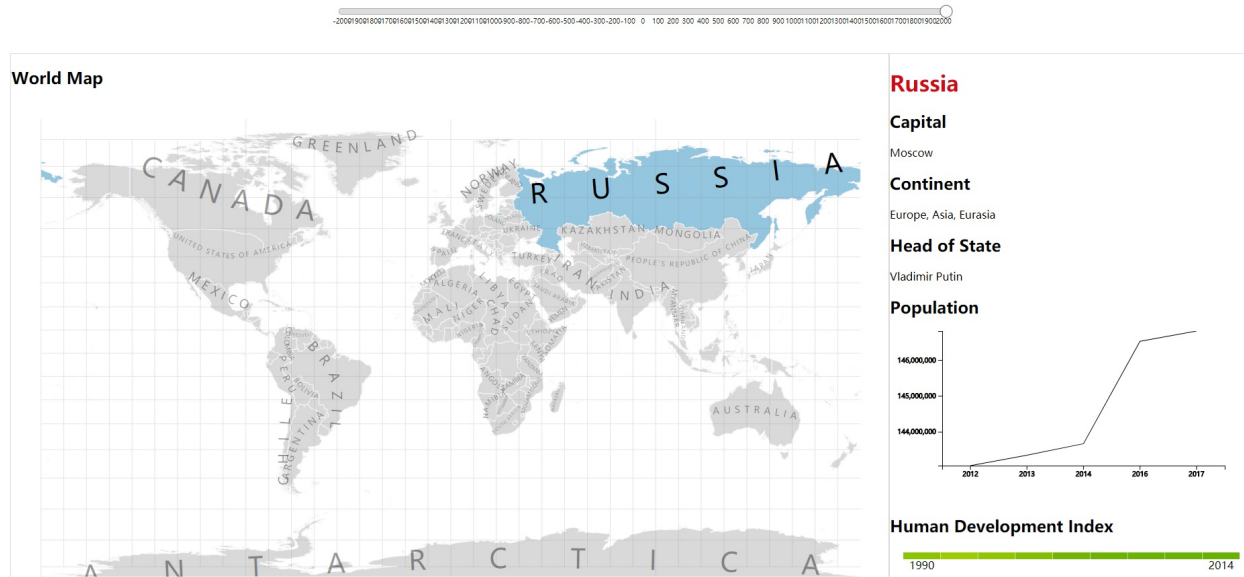


Figure 4: Current state of our project, Nov 10.