Data Domain:

I become interested in immigrants issues recently since it seems to become one of the most popular topics when people talk about policies made by Trump regime. To get a general idea of where do immigrants in the U.S come from, I searched for data from census.gov and found a dataset that contains the number of permanent immigrants to U.S by country of origin in every decade from 1950 to 2000.

Interaction techniques:

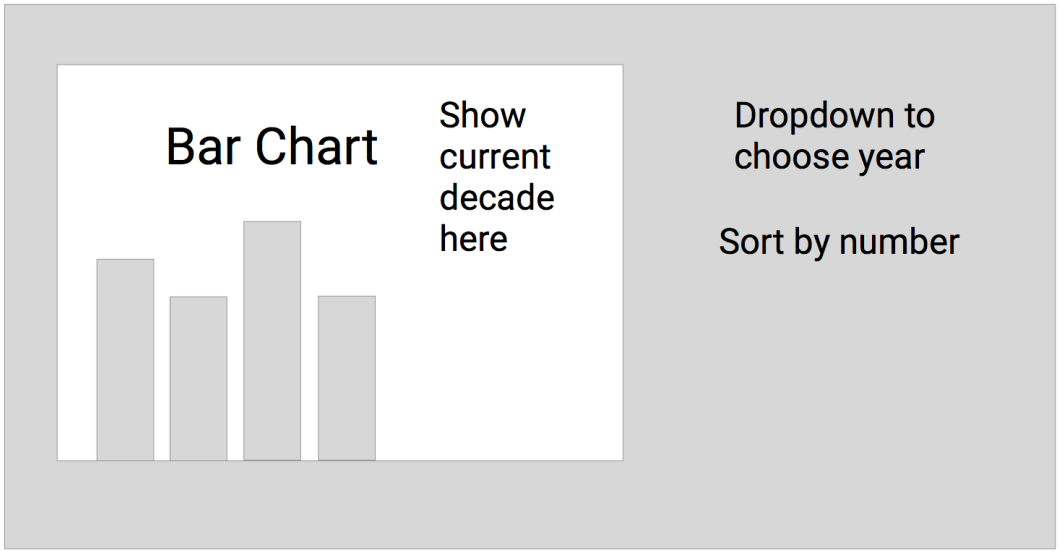
It's hard to show all data from five decades in one chart since there are more then 130 countries involved. I feel I'd better only show data for one of the decades between 1950-2000 initially, and use dynamic query to give users the option to explore data for different decades. Dynamic query could also help them navigate between different datasets to see the difference if they like.

The other techniques I thought about would be search, filter and sort. I think it would be helpful to let user search for some certain countries to see the results or let use sort the data to see the top ones. But it is not very usual to add filter to visualization. Thus I decided to use sort since ranking is also very important when there are more than one hundred data points. It helps to see the trend.

Tooltip is also needed to show the name of the country and detail of the data when the data point is hovered, it'll also be helpful to show the history data from all five decades to compare to the current one.

Application:

- Storyboard



Filled with color to show number

- Features

The first visualization is a bar chart that initially shows the data for 1990s (the latest data in the dataset). I want to add a dropdown menu to let use choose from different decades, and a sort function to let users view bar chart in different ways, either sorted by alphabetical order of Country name, or sorted by the number of immigrants in descending order. So users can easily find the country they interested in or get a sense of which countries are among the top ones. To clearly tell user what's the decade of data they currently viewing, I think it will be better to add a label on the top of the chart to indicate the decade.

The second chart is a world map. My dataset is about number of immigrants come from all over the world, I think there is no better way to show that data than using a world map. I can use color filling to show the number of immigrants and add hover function to let people know more detail.

I also want to add zoom in zoom out feature so people can view data for smaller countries easily and won't have a hard time clicking around to find the country.

- Concerns

There are too many countries to show, it's so hard to fit more than 100 countries into one single screen. I tried to find solutions for this problem, and come with solutions below:

  - Separate countries by continent
  - Separate by ranks (top50/bottom50)
  - Show top 50 for different decades

Another thing I noticed is that Mexico has huge number that when plotted with data from the other countries, it will be so hard to scale. For example, in 1990s, the number for Mexico is 9 million while the second place only has 1.3 million. I'd like to

try to adjust the scale and make the maximum domain half or even one third of the number of Mexico. This will make data for other countries more viewable, but it is also very deceptive since the length of the bar for Mexico should be longer.

I also met lots of problems since the data is from 1950 to 2000. During that time, lots of events that affect the territories and regimes of countries happened. The Soviet Union collapsed and became many smaller countries, Yugoslavia also collapsed and I don't know which country should I fit the data into. This situation made it harder for me to calculate and process the data. The world map might still be deceptive since some of the data is missing and some is not well processed.

Tool:

I use d3 to draw the bar chart and added a d3 library "d3.geomap" to draw the world map since d3 is the visualization tool that I'm most familiar with and it has great documentation and plenty of resources I can take advantage of. The d3.geomap library is easy to learn and I picked it up in 10mins. It is a great tool to visualize my dataset, which contains data for different countries. It actually provides only few customizable features, but it is not very hard for me to figure out its source code and customize it to get more features I need, thus I still decided to use it since it is very flexible.

Data source:

I spend around 10 hours in coding and clearing the data to get the visualizations properly rendered. I spend maybe 3 hours to figure out which is the appropriate way to visualize my dataset and choose interaction techniques. I'd say debugging while adding interaction components took me the most time, but clearing the data might be the same. The csv file I download from census.gov contains lots of spaces, commas, and slash for indentations. The data also contains the sum for regions and every continent. It also has data for some countries that are not exist any more. Also the country name is not in standard Alpha-2 or Alpha-3 format, it even contains parenthesis with annotation in it. I spend lots of time convert country name to Alpha-3 code.

Data source URL:
https://www.census.gov/population/www/documentation/twps0081/twps0081.html