Data Visualization Final Project Proposal

1.Basic Information

Project title: Population Migration in motion of USA between 2004 and 2015
Team: Zheng Nie, znie@wpi.edu, nzholmes;
Shijian Li, sli8@wpi.edu, lawdpls;
Yi Jiang, yjiang@wpi.edu, YiJiang93;
Repository:https://github.com/YiJiang93/Population-Migration-of-USA

2.Background and Motivation

As we search the population migration visualization examples, we could find the majority type of this visualization focus on statistical description visualization and most visualization result are stationary. They just use statistical numbers to reflect the population migration without showing the trend.

We are wondering how to indicate the the location change directly in the map so that readers can easily see the migration process clearly. Later, we found a visualization technique called flow visualization, which is used to make the flow patterns visible, in order to get qualitative or quantitative information on them.

Generally speaking, flow visualization is the art of making transparent flow(air,water,etc) patterns visible because their flow patterns are invisible to the naked eye without methods to make them visible. Here we consider flow visualization as an approach to reflect the flowing trend and try to take advantage of this technique to visualize population's flowing variation.

3.Project Objectives

We decide to adopt the design of flow visualization because its characteristic of being fluid perfectly depicts the migration pattern and makes big improvements on existing static visualization in migration.

A hard challenge as it is, the "pain" it inflicts on us will undoubtedly be a great learning experience in which we can gain abundant treasure.

The guiding questions that directs us at this stage are as follows:

1. The US population migration trend in motion between 2004 and 2015
2. The population migration trend of a specific state over a time period
3. The regional population trend by factors such as sex, age, race, income

The aspects we are focusing on are:

1. How to simulate flow processes using d3 or other JS frameworks?
2. How to make reasonable arrangements between dynamic charts and static charts in the visualization project? For example, given a US migration flow visualization, is the bar chart that depicts trend for a state still needed?
3. How to present more information using visual channels or charts as simple and clear as possible? In other words, how to make a visualization carry as much information as possible?
4. The last aspect lies in the code abstraction. How we implement the flow design such that it can be reused in other similar cases?

4. Data

From where and how are you collecting your data? If appropriate, provide a link to your data sources.

Our data source is the census of governments, which provides state-to-state and regional migration flows over years.

http://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-to-state-migration.html

http://www.census.gov/topics/population/migration/data/tables.All.html

5.Data Processing

The datasets listed above contain the population migration information of every past year, each of which is represented in a .xls spreadsheet file. There are in total two categories of data that we want to utilize, namely the State-to-state Migration Flow data and the Migration/Geographic Mobility data.

Each data file has the time scale of 1 year, which means all dataset is annual and does not get carried over. In order to get the trend of population migration, we decide to set the scope to 10 years. Typically the idea would be web crawling the Census website. That said, in our case we scrap this idea and do it manually because:

1. The datasets are not real-time. As a matter of fact, next update will happen way after this semester.
2. Each file is about 1/10 of the whole temporal scope, it would take no efforts to download it manually.

For State-to-state migration Flow data, the table contains 50 states as rows and columns, plus extra columns to depict what the movement is, e.g. stay in the same household, stay in the same state or move to a different state. Every data entry has two values, an estimation value and a margin of error. For our visualization, we decide to use estimation value and discard the margin of error because according to observation, all entries has similar margin of error. In general, we'll be using a 52x54 table for each year.

For Migration/Geographic Mobility data, there are more data variables to access. The Census website provides tens of files for every year, including different categories like sex, age, race, ethnicity, employment status etc., on different subjects such as general mobility, migration flow, inmigrants/outmigrants net gain and reason of movement etc. We use only the migration flow for our project because it can demonstrate our design goal the best. We'll be using the different categories to deepen the data layers and varieties for the regional charts. In general, for each file we have 12 columns(four regions each corresponds to other three) and 61 rows(nine categories). The categories are shown as below:

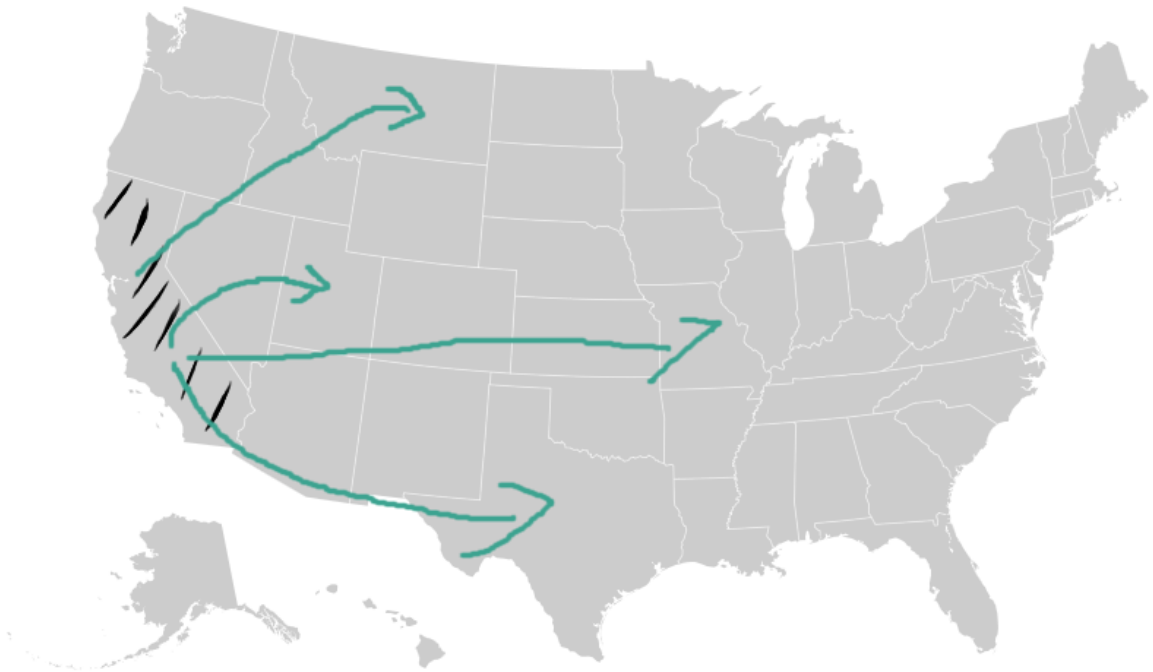
1. Sex
2. Age
3. Race and Hispanic Origin
4. Relationship to householder
5. Educational attainment
6. Marital status
7. Nativity
8. Tenure
9. Poverty Status

For now we decide to process the downloaded data by writing a script using SheetJS. We might change to do it manually since the data files are not in too large quantity and are highly organised.

6.Visualization Design

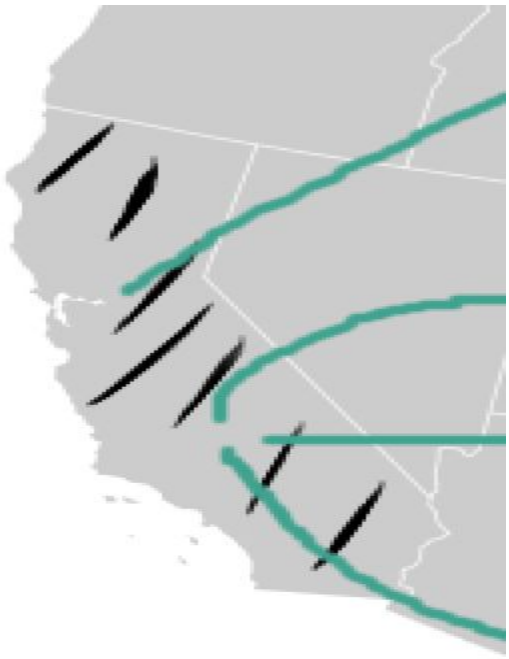
Here are some results of us using the Five Design Sheet method to design the layouts and visualization types of our project.

Vis 1:



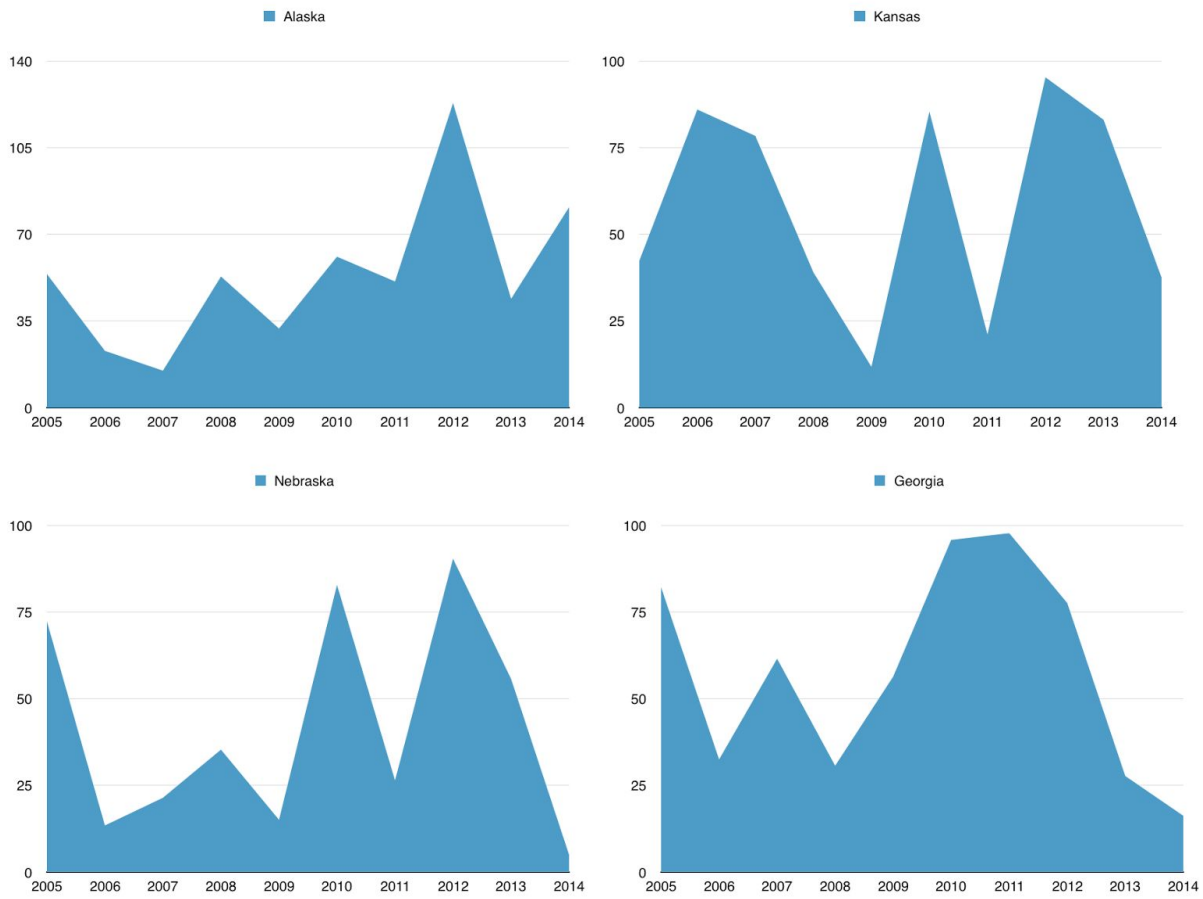
The state-level flow map.

For this map, our design is to have a US map with interactivity that we can select each state, and have a steady flow represents the outmigrants of this state to other states.



The focus of this visualization is to have each state selectable, and after that the migration flow must automatically generate. Deselect this state will terminate the flow.

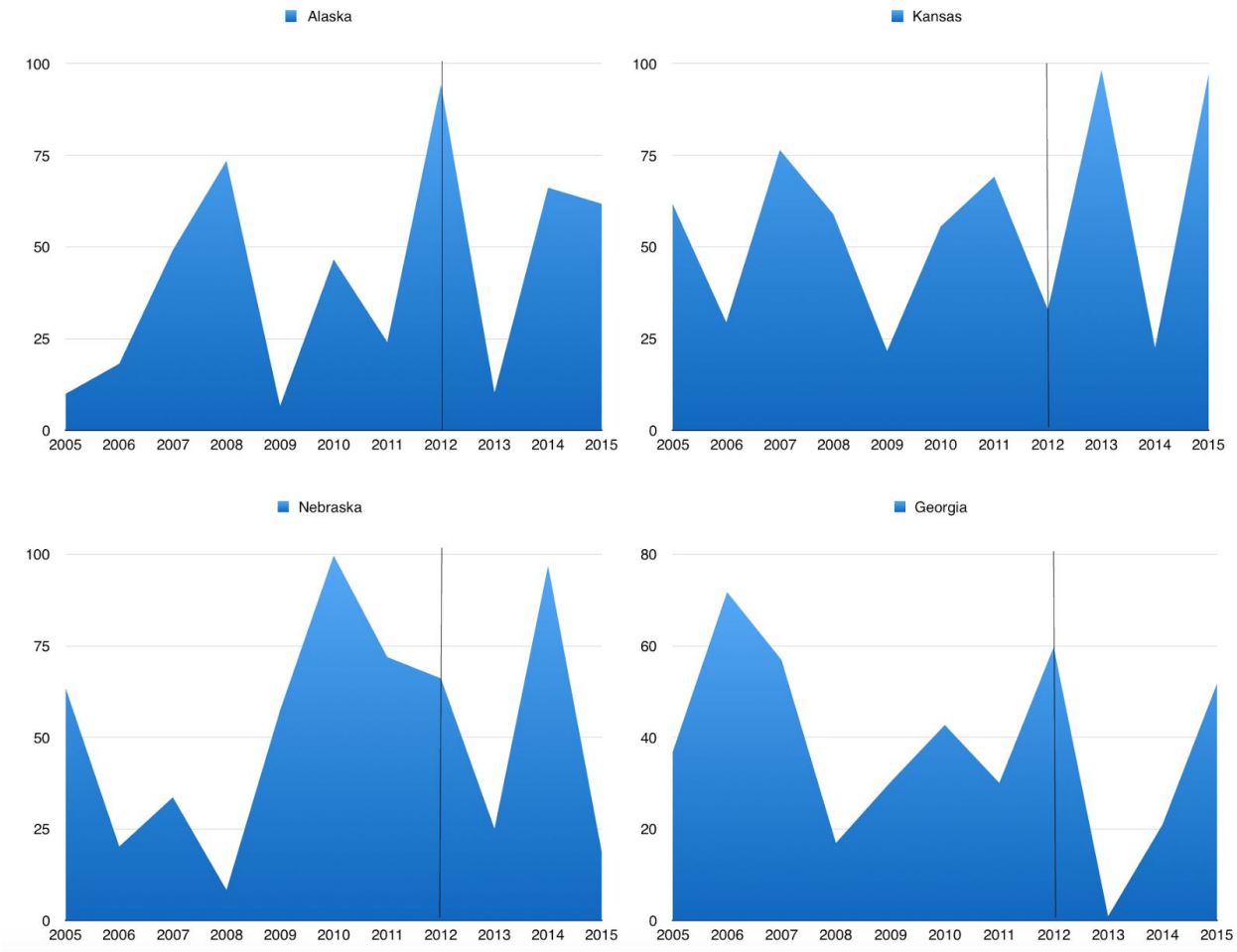
Vis 2:



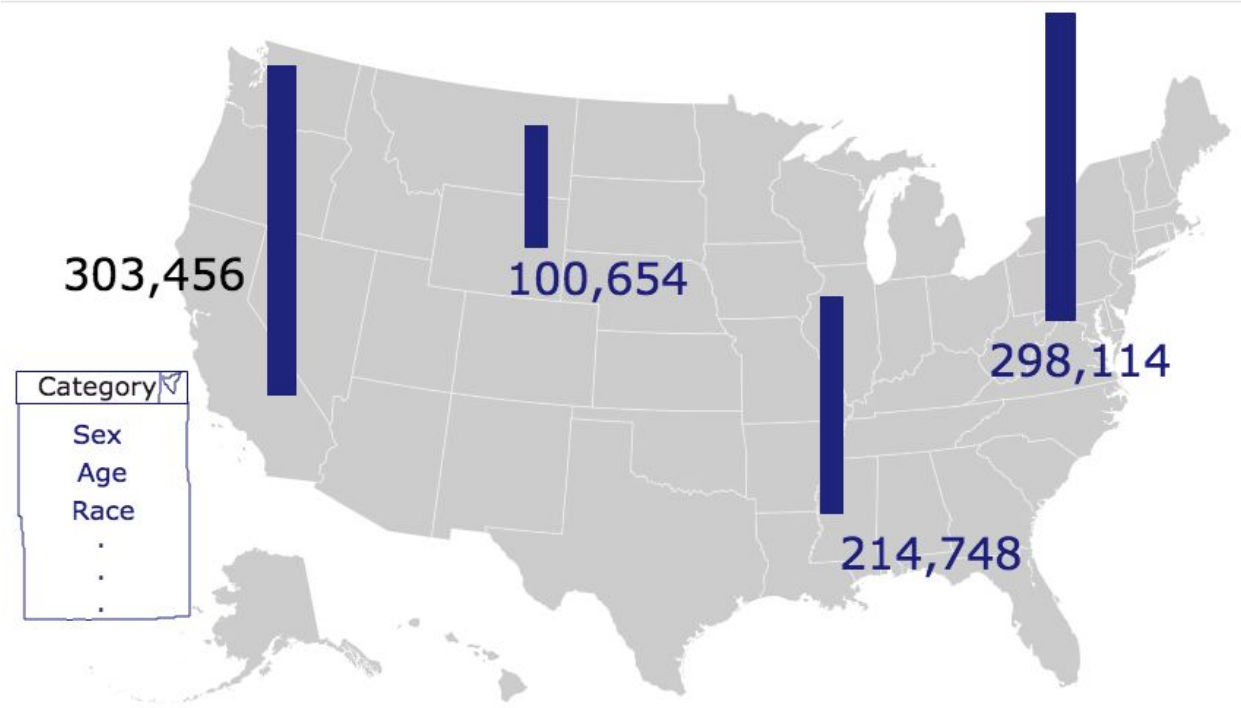
A matrix area chart

We want to include this chart because it can clearly contrast and compare the in/out flow of each state. Here we plan to display all 50 states at once, and have mouse interactivity to select a year of all states in a synchronized fashion.

The focus of this visualization is to display the same year's data for every state. Show below as the black line. Here we first select "inmigrants number", then mouseover the year of 2012, this should give us the respective inmigrants number(maybe shown under the X axis) of all states in 2012.


Alaska
100
75
50
25
0
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
Kansas
100
75
50
25
0
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
Nebraska
100
75
50
25
0
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
Georgia
80
60
40
20
0
2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015

Vis 3:



The regional map

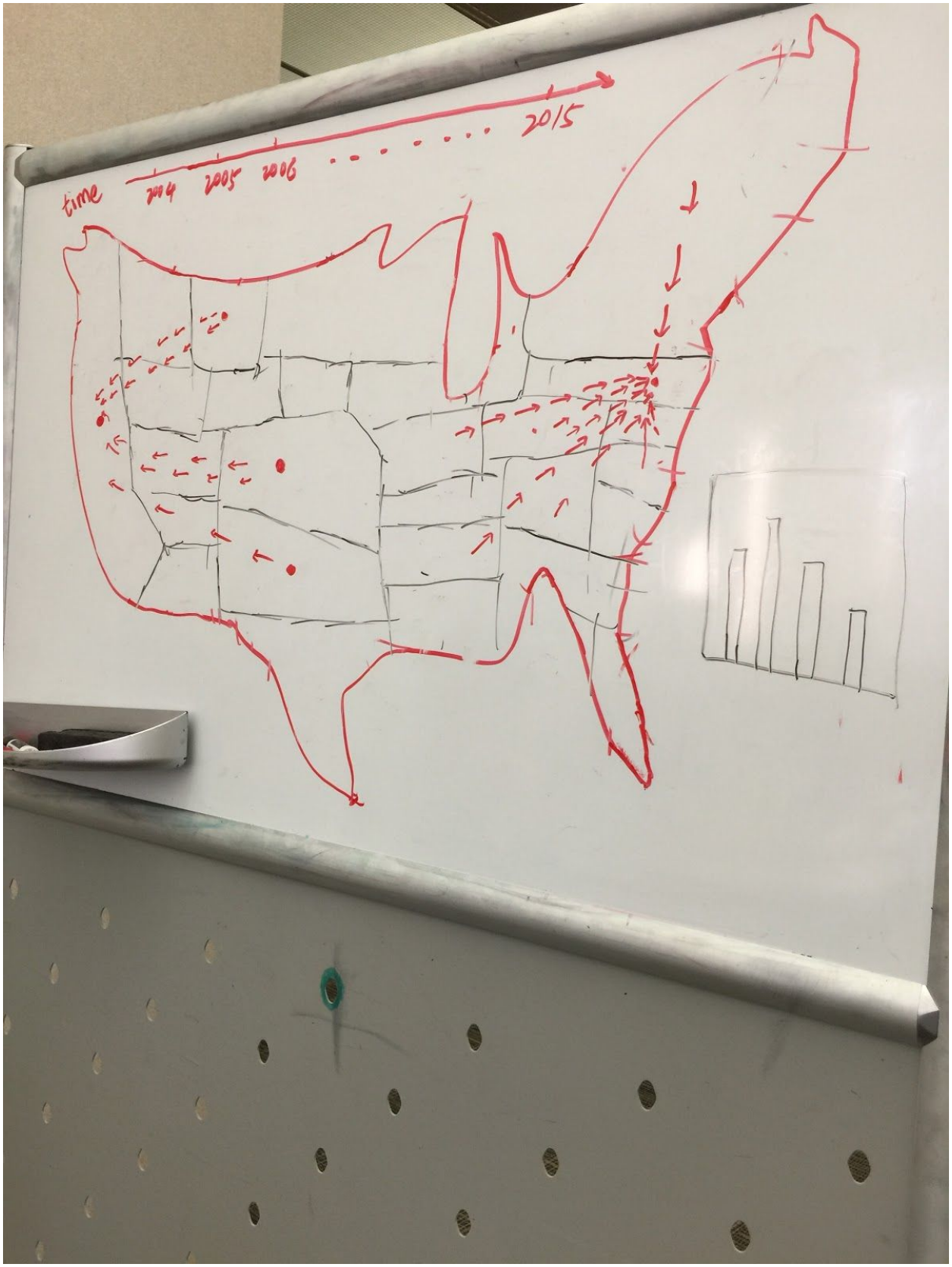
Yet another map visualization. In order to make this map different from the state map, we decide to emphasize more on the statistics aspect of it. We'll have many categories to filter through the data, as well as some other interactivity to spice it up.

The focus of this map is the ability to change between different categories and how well we can integrate them together.

Final Design:

As we feel a single visualization won't be enough to capture what we meant to express, we decide to adopt all above designs. Thus basically we'll have several pages to switch between, each page displaying a chart that we have designed.

Also we decide to make change to the State-to-state flow map. As a flamboyant map, have only one state selectable at a time really doesn't do justice to its nature. We want to display the flow of all states at once, with a timeline to automate the progression of time.



The revised State-to-state map

Our final design would be this map plus the 2nd and 3rd vis from above.

7.Must-Have Features

This visualization is an interactive map including the migration from states to states and from regions to regions. These features should be contained:
As for state to state migration map:

- Moving timeline (automated moving or can be specified by users)
- Flowing trend based on states' data
- User can select the immigration state or outmigration state
- Mouseover function
- Static charts

As for region to region migration map:

- The map of USA should be divided to 4 part:Northeast,Midwest, South and West
- Mouseover function
- Pull down list to see migrants' information like sex,age,marital status,etc
- Basic charts to show statistical analysis

8.Optional Features

- Flowing trend based on regional data
- Migrants' income information visualization based on another table file from Census website

9.Project Schedule

- 1st week (10.31-11.7). Collecting, cleaning and preprocessing state and regional data.
- 2nd week (11.8-11.14). Making state-to-state migration flow visualization.
- 3rd week (11.15-11.21). Add static charts like bar chart for each state for the purpose of analysis.
- 4th week (11.22-11.29) Prototype presentation and add analytical static charts for regions.
- 5th week (11.30-12.5) Write process book and construct project website.
- 6th week (12.6-12.12)Strength our functionality and create a two-minute screen-cast.

Reference:

Region Definition:
http://www.census.gov/topics/population/migration/data/tables.All.html
Region Migration:
http://www.census.gov/econ/census/help/geography/regions_and_divisions.html
State to State Migration:
http://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-to-state-migration.html