CS107 FINAL PROJECT PROPOSAL

UNITED STATES OF AMERICA CENSUS VISUALIZATION

Simon Rouben Malian
SM in Computational Science and Engineering Candidate
Harvard School of Engineering and Applied Sciences

BACKGROUND AND MOTIVATION

Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

As a graduate student who is in his second semester of studies at Harvard, I am interested in learning more about the cultural and socio-economic factors of America. I came to the conclusion that US Census data will be a great way to showcase multiple layers of information that are linked with one another – e.g. the population's gender ratio at a country level, state level, municipality level.

The visualization will give me a good opportunity to learn more about America while at the same time allowing me to experiment with different visualization types that all work cohesively to communicate information – e.g. a choropleth map which takes center stage and is supported by other visualizations which provide further detail (e.g. a tree map, stacked bar graph).

The work can be used by others – e.g. public policy students – to make informed policy, planning and funding decisions.

PROJECT OBJECTIVES

Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.

This project aims to reveal whether there are any correlations between location, socio-economic distribution, ethnic origin, household income, and other parameters I may add in the event they appear useful during the course of coding.

The benefits of understanding the current population characteristics will influence user's understanding of each location. For public policy students who may use the visualization: it will help facilitate developing realistic projections regarding further

trends and challenges (e.g. aging population) that may be utilized, neutralized or reduced through proper planning.

DATA

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

The data for this analysis has been obtained from the US Census Bureau. The Bureau gathers up-to-date population statistics from decennial censuses and annual surveys further producing population estimates with analysis of the demographic components of population change for states, counties, cities, and towns at the national, state, and county level. It also provides population projections that illustrate various possible courses of population change based on assumptions such as future birth and death rate, domestic and international migration.

A copy of the intended dataset has been made available within the project's Github repository.

DATA PROCESSING

Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

The project will hopefully only require light data cleanup. The main units of the data will include the population amount distributed by zipcode. The analyzed data encompasses various core components such population ethnic diversity, male/female population ratio, median age, income per household, and house value.

VISUALIZATION

How will you display your data? Provide some general ideas that you have for the visualization design. Include sketches of your design.

The main component of visualization will be in the form of a map (choropleth), showing distributions of different data attributes. I will also consider introducing control elements to allow interactive exploration of the data (e.g. show the data for zipcode to analyze whether it is above/below certain threshold).

Visualization sketch appears on next page

Ethnicity distribution

Business measure by area type

Treemap population

Gender by state
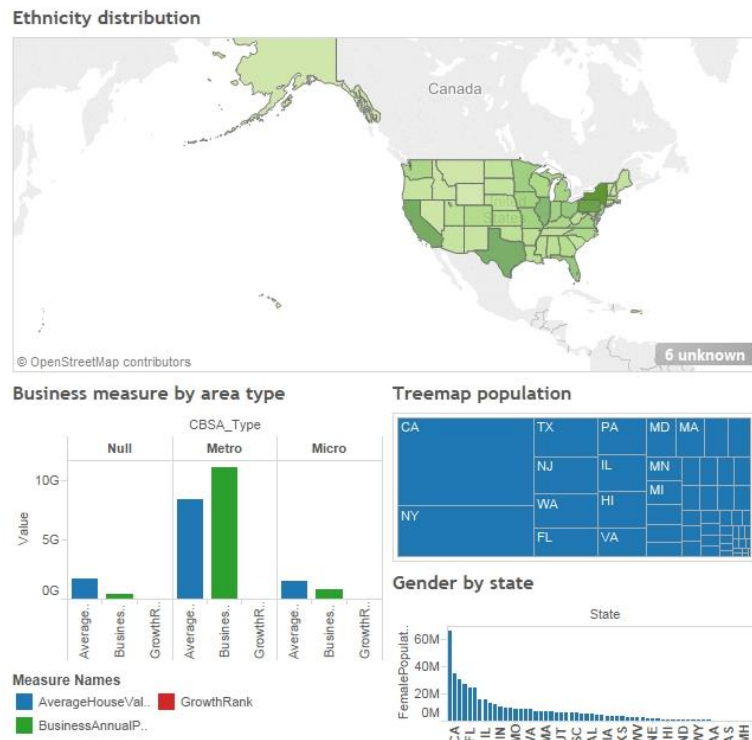
MUST-HAVE FEATURES

These are features without which you would consider your project to be a failure.

The core feature of the project is a population map that illustrates the distribution of main socio-demographic parameters (ethnic group, income level etc.) by state (or probably a more granular level, such as counties).

A choropleth map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as population density or per-capita income. The choropleth map provides an easy way to visualize how a measurement varies across a geographic area or it shows the level of variability within a region.

To enable interactive layer, I aim to introduce control elements, which allow the user to filter, select subsets of data across dimensions (such as ethnicity, gender, income etc.) on the map.

I consider those to be a must-have features, as map layout and selector control elements allow the user to extract and present geo-location data in the most possible natural way – i.e. on the map

OPTIONAL FEATURES

Those features which you consider would be nice to have, but not critical.

The project will be supported by complimentary charts showing yearly trends in the population dynamic, as well as show general charts with ethnic / income distribution for the whole US population, which will helps to get a brief information about population attributes across the country.

In the event I finish things early and have spare time: I may try to add additional elements – e.g. number of attractions in a given area, etc. – to make the material more interesting.

PROJECT SCHEDULE

Make sure that you plan your work so that you can avoid a big rush right before the final deadline. Write this in terms of weekly deadlines.

My intended project schedule is as follows:

Submit proposal and create project repository	**April 03 2015**
Finalize data and intended design	April 10 2015
Finish coding and begin final testing prior to submission	April 15 2015
Submit Milestone 1	**April 17 2015**
Obtain feedback from others and make updates	April 19 2015
Individual project review with TF	**April 21 2015**
Make updates based on TF feedback	April 28 2015
Finish coding and begin final testing prior to submission	May 03 2015
Submit final project	**May 05 2015**
Present at Harvard SEAS Design Fair	May 06 2015