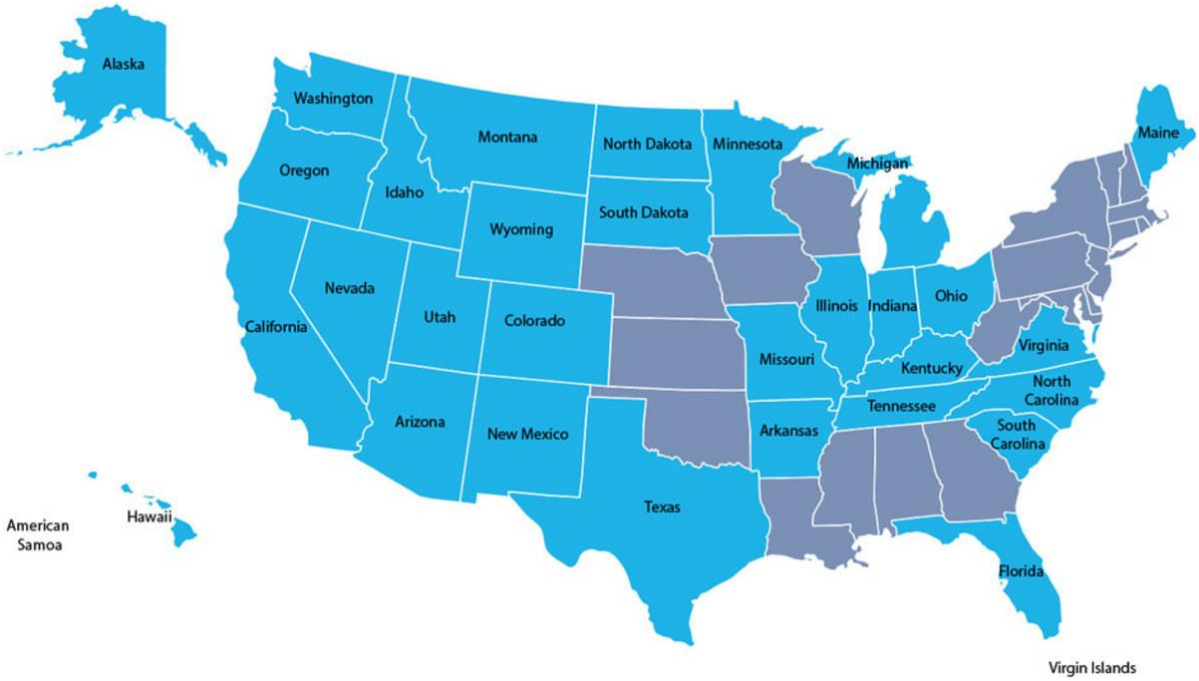
CSE 564 VISUALIZATION AND VISUAL ANALYTICS

UNITED STATES DEMOGRAPHICS - FINAL PROJECT REPORT

PRESENTED BY:

KONDETI MUKESH GOUD (112689870)



INTRODUCTION

The Goal of this project is to build a dashboard to visualize the changes in American society and demographics over time and also, with the help of the visualization, analyze the effects of these changes on various other factors of the society to produce interesting, statistically significant results.

DATASET

For this project, I've considered United States census data of the last five decades. They were downloaded from Kaggle and after downloading, cleaning of the dataset like: removing NaN values, removing outliers using basic plots is performed. After the cleaning of data is completed, then the five datasets are clubbed into a single dataset and then it is used for rendering.

This final dataset consists of 255 rows and 36 attributes and it was downloaded from Kaggle. 14 attributes were added to the dataset which were derived as combination of initial attributes Attributes of the dataset : 'STATE', 'Total_Population', 'Urban', 'Suburban', 'Rural', 'Male', 'Female', 'Native', 'Foreign', 'Europe', 'Asia', 'Africa', 'Oceania', 'Americas', 'MedianIncome', 'PerCapitaIncome', 'Poverty', 'HousingUnits', 'Male_Ratio', 'Female_Ratio', 'Sex_Ratio', 'Urban_Ratio', 'Suburban_Ratio', 'Rural_Ratio', 'Location_Ratio', 'W_Ratio', 'Native_Ratio', 'Foreign_Ratio', 'Origin_Ratio', 'Europe_Ratio', 'Asia_Ratio', 'Africa_Ratio', 'Oceania_Ratio', 'Americas_Ratio', 'Poverty_Ratio', 'Housing_Ratio', 'Male_5_Ratio',

'Male_8_Ratio', 'Male_11_Ratio', 'Male_A_Ratio', 'Male_B_Ratio', 'Male_G_Ratio', 'Female_5_Ratio', 'Female_8_Ratio', 'Female_11_Ratio', 'Female_A_Ratio','Female_B_Ratio', 'Female_G_Ratio', 'Educated_Sex_Ratio', 'College_Educated_Ratio', ‘year’.

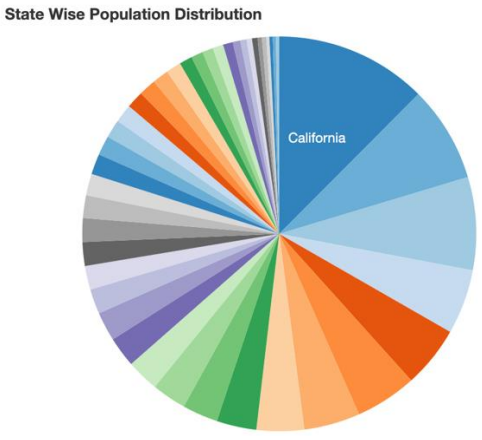
PROBLEM STATEMENT

Society is constantly evolving, and we want to develop a way to visualize this evolution through time so as to make it easier for policy makers to heed to the changing needs of their county, state and county. The goal is to visualize the data in a way that pinpoints and brings forth the ways the demographic changes affects the economics and other areas of the society and to show that these are important factors that need to be considered when designing policies and governance.

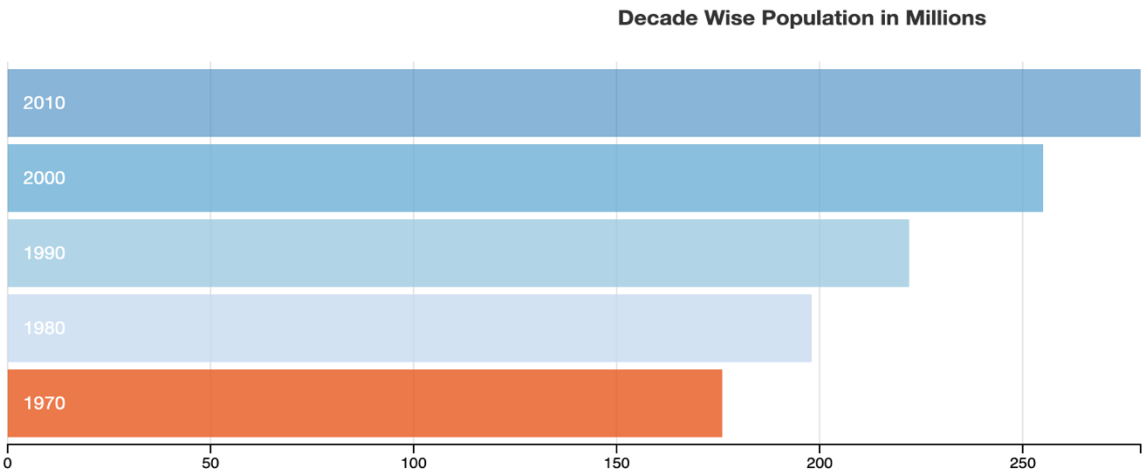
DASHBOARD DESIGN

The structure of the dashboard consists of six parts:

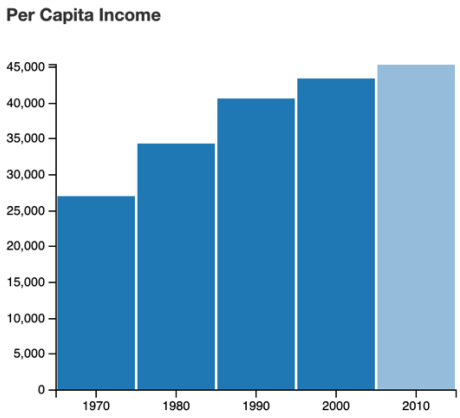
1. A Pie Diagram to represent various individual states based on the cumulative population over the past five decades.



2. A Row chart between years and population: This chart represents the population growth in the individual decades of a particular state when the state is clicked on the pie diagram. By default, this chart represents the population growth of entire United States.

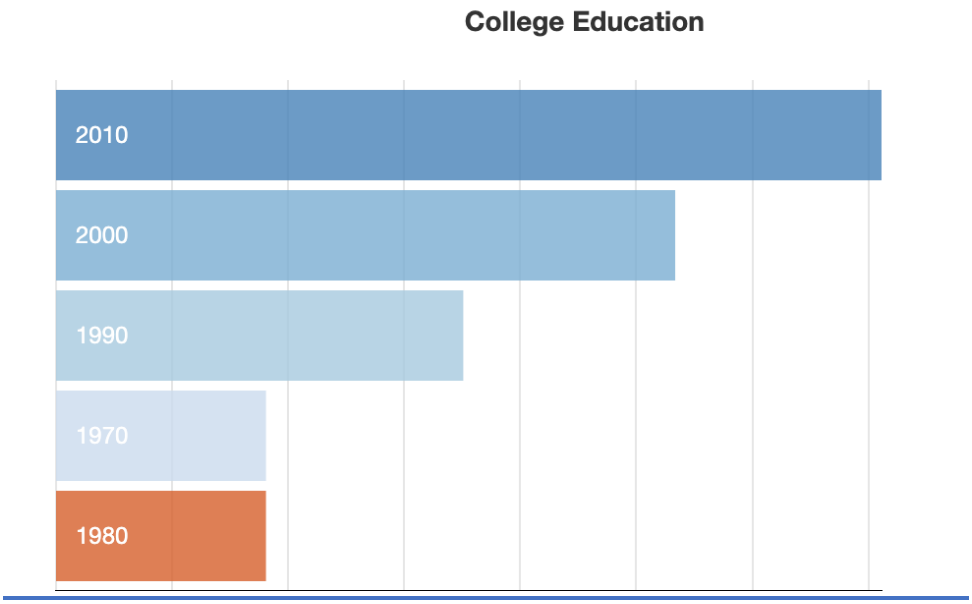


3. Bar diagrams of Per Capita Income verses years: This horizontal bar chart represents the growth of per capita income over the years of the entire United States. State wise per capita income can be viewed by selecting the particular state from pie chart.



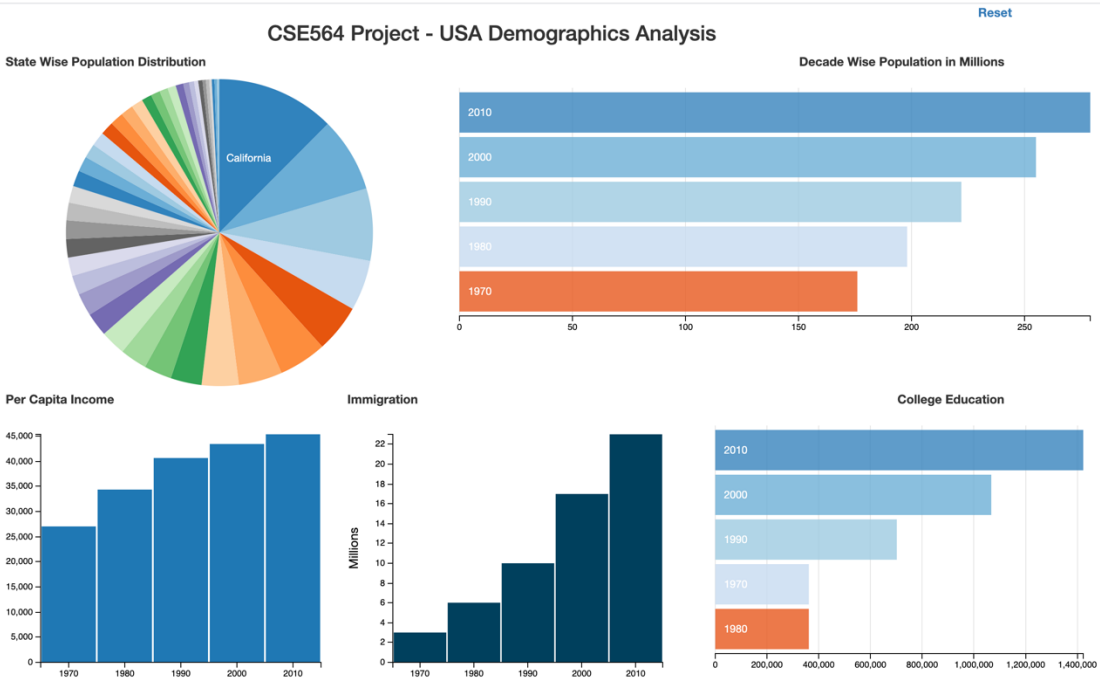
4. Bar diagrams of Immigration verses years: This bat diagram represents the amount of immigration which has happened over the past few decades of the entire United States.

5. A row chart to represent the immigrants pursuing education with respect to years: This chart represents the number of people who got educated in the United States over the period of the last five decades.



6. A Reset button to re-render all the visualizations. This button is located at the top right part of the dashboard.

INTERACTIVITY OF THE DASHBOARD



The above image represents the entire information of United States demographics at a glance. The dashboard is interactive i.e. clicking on any part of the dashboard changes the other state of the dashboard and vice-versa. By selecting a few parameters from the dashboard, we can view the information we need and ignore the irrelevant information. Reset button is also provided to re-render the charts from any point.

INTERESTING INSIGHTS

1. **Immigration vs State**: From the above dashboard, if we select only the most populous states (states with large part of the pie chart), we can see that the entire immigration bar plot is filled with bars (i.e. entire immigration happens only in these populous states) and if we select the less populous states, we

can see that no immigration happens in these states. Thus, we can conclude that immigrants are interested to live in most populous cities.

2. **Reduced Growth of Per Capita Income**: In the last two decades, the growth of per capita income is less when compared to the per capita income growth in the first two decades in the graph. At the same time, if we observe the immigration levels for the last two decades, it has gone considerably high. This situation can be attributed to the increased immigration made more supply and thus reducing the demand.

3. **Immigration and College Education**: There is a upward curve for immigrants entering United States with an interest in pursuing education. This trend is evident in the bar plot show in the dashboard. This can be attributed to many reasons like: education quality, recognition of degree and luring job market. The first two decades witnessed less immigration and thus less interest in pursuing education among immigrants.

CONCLUSION

In conclusion, I am hopeful that this study on United States demographics can help everyone with some interesting analysis and some meaningful, statistically significant results that can be used in the real world. This visualization can be applied to real time usage to help with governance and bringing to forefront counties/states that are failing to provide a good economic, educational and healthcare opportunities to its people.

REFERENCES

1. https://www.kaggle.com/muonneutrino/us-census-demographic-data
2. https://usa.ipums.org/usa/
3. https://github.com/dc-js/dc.js
4. http://www.d3noob.org/2013/08/add-pie-chart-in-dcjs.html
5. http://bl.ocks.org/d3noob/6584483