Interactive Visualization of Latino Political Participation in Nebraska and USA

Farhana Luna* Xiaoyue Cheng† Mahbubul Majumder‡

Abstract
How would the Latino population impact on the future regional and national elections in Nebraska and USA? In this research an interactive graphical tool is developed to visually analyze various data sources from the US Census Bureau and University of Nebraska database. The tool is built in an online Shiny Dashboard so users will conveniently reveal patterns in Latino population growth, track Hispanic political participation, and compare Latino people with other races. Data access and data cleaning methods will also be introduced in the paper.

Key Words: Latino population, data visualization, shiny dashboard, interactive graphics, census data.

1. Introduction

Hispanic Americans constitute 17.6% (56.6 million out of 321.4 million) of United States population in 2015, estimated by United States Census Bureau . As a second fastest-growing ethnic group, Hispanics are playing a more important role in American society and politics. A large amount of data about Hispanic population were released for free each year by United States Census Bureau [1] and other organizations, but not much information was dug or properly used. One possible reason is the lack of technique skills to clean raw data and link multiple data sources. Another difficulty is how to discover useful information and deep thought from the routine data. This research intends to provide advanced visual representation of data graphics that would be helpful to explore and understand the inside scenario of Hispanic population data.

This research started with a need from the Office of Latino/Latin American Studies (OLLAS) in University of Nebraska at Omaha (UNO). The office develops policy-oriented and community-relevant research for Latino American in the Nebraska and Great Plains region. As the amount of data increased year by year, researchers from OLLAS realized that data visualization will strength their research and keep the Latino study attractive to the community. But how to product good visualizations from unorganized data had become a big challenge. Therefore, we developed an interactive visualization tool that can temporally and geographically explore patterns in the growth of Latino population. Our research not only met the needs for Latino study, but also can be broadened to other social and political study which uses census data.

Section 2 introduces different data sources, issues and mistakes in the data, and our processing procedure. Section 3 introduces functionalists of the interactive graphical tool. Section 4 shows examples of real data analysis from the tool.

*Graduate student concentrated in Data Science, Department of Mathematics, University of Nebraska at Omaha, 6001 Dodge Street, Omaha, NE 68182. Email: fluna@unomaha.edu

†Assistant professor, Department of Mathematics, University of Nebraska at Omaha, 6001 Dodge Street, Omaha, NE 68182. Email: xycheng@unomaha.edu

‡Assistant professor, Department of Mathematics, University of Nebraska at Omaha, 6001 Dodge Street, Omaha, NE 68182. Email: mmajumder@unomaha.edu

[1] `http://factfinder.census.gov`

2. Data

2.1 Data sources

Five data sources were used for the research:

- Reported Voting and Registration by Race and Hispanic Origin, for States, by US Census Bureau[2]. This data resource has the longest history: it has been published every two years since 1980. But it is also the most problematic resource. We'll address the problems in Section 2.2.
- Citizen Voting Age Group, by US Census Bureau[3]. This data provided a 5-year population estimate by race in different geographies (nation, state, county, place, Minor Civil Division, Tract, and Block Groups). Eight variables are given: total population, total adult, total citizen, total citizen adult, and their margin of errors.
- Cartographic Boundary Shape files, by US Census Bureau[4]. To make use of the Citizen Voting Age Group data in different geographies, we have to collect the geographical information.
- Nebraska Hispanic Population by Sex and Age, provided by OLLAS. This data set has Hispanic population by sex, in ages $0, 1, \cdots, 84, 85+$, annually from 1990 to 2013.
- Nebraska Hispanic Population by county, provided by OLLAS. The data were collected in four years: 1990, 2000, 2010, and 2013.

2.2 Data challenges

As we mentioned above, the Voting and Registration data from US Census Bureau are problematic. There are five major challenges for using the data.

- Only scanned pdf files were provided before 1996. OLLAS provided us with data in Excel format but in the files there were many mistakes, for example, columns are misaligned, spaces and blank rows appears everywhere without any pattern. Figure 1 shows some typical typos: Numbers 0, 1, and 5 may be recognized as O, I, and S. GEORGIA was recognized as GI!DRGIA. Three rows may be recorded in one row. In addition, there were two table panels on the same page to save space, as shown in Figure 2.
- Some terms changed for a few times over years. For example, before 2000, "standard error" was provided along with percentages, during 2000-2002, "(90%) confidence interval" replaced "standard error", after 2002, "margin of error" was utilized. Another example is the race names. "Hispanic origin" was called "Spanish origin" in 1980-1984, "Hispanic Origin" in 1988-1996, "Hispanic" in 1986, 1998, 2000, and "Hispanic (of any race)" since 2002.
- Variables in the table were not consistent. Generally speaking, more details were included as time went by. In 1980, there were only 3 races and 5 variables: total population, percent registered and percent voted with their standard errors, but starting from 2006, there were 8 races and 12 variables including the previous variables, plus total citizen, total registered, total voted, percent registered citizen, percent voted citizen with their margin or errors.

[2] https://www.census.gov/topics/public-sector/voting/data/tables.html
[3] https://www.census.gov/rdo/data/voting_age_population_by_citizenship_and_race_cvap.html
[4] https://www.census.gov/geo/maps-data/data/cbf/cbf_blkgrp.html

GI!DRGIA					
Total	3,558		65.6		1.3
White	80.6		67		1.5
Black	19. I		59.8		3.8
MINNESOTA					
Total	2, 916		83.7		1.1
Percent	100		(X)		(X)
White	97. 9		83.8		1.2
MISSISSI PPI					
Total	1,602		81. I		1.4
Percent	100.0		(X)		(X)
White	69. 5		85.2		1.5
Black	29.8		72.2		3.4

Figure 1: Example of typos in the Excel data converted from pdf files.

- Data tables in 1980s have a mixture of total and percent measures in the same column. Unlike a regular data table, which records same variables for different rows, several tables in early years put the percentage of White, Black and Hispanic people under the total population, as shown in Figure 2. This issue was not realized until plots were made from the dashboard, which showed that the White population was trivial in 1980s but suddenly exploded in 1990s.

State, race, and Spanish origin	All persons1	Reported registered		Reported voted		State, race, and Spanish origin	All persons1	Reported registered		Reported voted	
		Percent	Standard error	Percent	Standard error			Percent	Standard error	Percent	Standard error
UNITED STATES						HAWAII					
Total	157, 085	66.9	0.2	59.2	0.2	Total	601	61.9	2.0	55.9	2.1
Percent	100.0	(X)	(X)	(X)	(X)	Percent	100.0	(X)	(X)	(X)	(X)
White	87.6	68.4	0.2	60.9	0.2	White	32.4	66.5	3.5	59.0	3.6
Black	10.5	60.0	0.7	50.5	0. 7						
Spanish origin	5.2	36.3	1.5	29.9	1.4	IDAHO					
						Total	613	73.5	1.6	70. I	1. 7
ALABAMA						Percent	100.0	(X)	(X)	(X)	(X)
Total	2,617	70.2	1.5	56.4	1.6	White	98.6	73.6	1.6	70.2	1.7
Percent	100.0	(X)	(X)	(X)	(X)						
White	73.5	73.3	1.7	59.2	1.8						
Black	26.2	62.2	3.7	48.9	3.8	ILLINOIS					
						Total	7,885	72.8	0.8	66.4	0.9
ALASKA						Percent	100.0	(X)	(X)	(X)	(X)
Total	244	66.6	1.4	56.8	1.4	White	86.4	74.0	0.9	67. 7	0.9
Percent	100.0	(X)	(X)	(X)	(X)	Black	11.9	72.I	2.9	64.5	3.1
White	82.9	69. 7	1.5	60.9	1.6	Spanish origin	4.0	41. 5	7.8	38.2	7. 7

Figure 2: The first column "All persons" mixes the total population and race percentages.

- Missing data exist. The symbol (B) means that the base is less than 75,000 and therefore too small to show the derived measure. Figure 3 shows a few (B)'s for ethnic minority groups of Nebraska in 2000. We should keep alerted when the percentage is close to 0 or 1. But when it is around 0.5, like the Hispanic percent citizen in the figure, it is not a problem to calculate the percentage, as long as the margin or error is conservative.

NEBRASKA	Total	1,205	1,176	97.6	0.8
	Male	579	567	97.9	1.1
	Female	626	609	97.3	1.2
	N-H White	1,078	1,075	99.7	0.3
	N-H Black	46	46	(B)	(B)
	API	14	13	(B)	(B)
	Hispanic	56	32	(B)	(B)
	White	1,133	1,106	97.6	0.9
	Black	46	46	(B)	(B)

Figure 3: Example of missing data in the percent estimate where the base is small.

2.3 Data cleaning

Given all the challenges above, cleaning the data and replacing missing values was not simple. Some initial pre-processing was performed in Excel, like aligning the columns, and moving the right table panel to the bottom of the left panel. After that, data cleaning, reshaping, reordering and other additional necessary processing were all performed in **R**.

In data processing, we found out and removed all the typos, removed extra symbols and letters like dollar sign or percentage sign, unified the variable and race names over years, split the mixed column into two columns for totals and percentages, recalculated the missings (B) from existing data, matched the letter cases between data files and geographical information, reordered data for creating maps, melted and cast the data tables into a usable structure, combined data from different resources.

After cleaning the data, we started to explore and analyze the data using many graph types like bar chart, dot plot, time series plot, and maps, in order to find the most effective visualization and interaction that fit the data.

3. Interactive Graphical Tool

Nowadays data visualization is an important part of data science, and it serves many different purposes. Visualization is useful for understanding the general structure and patterns of the data. When analyzing large and complex data sets, static plots, which do not allow interacting with graphics, cannot present all information of data effectively. On the other hand, computer-user interaction gives flexibility and control to users. For example, users can view the inner status of data from different directions by rotating an interactive graphics. Interactivity has become more imperative in exploratory data analysis. For this reason, we developed a dashboard using the **R** packages **shiny** and **shinydashboard**. The interactivities were supported by packages **ggvis**, **plotly**, and **leaflet**.

3.1 Functionality

The dashboard was designed with a uniform style but seven different tabs for seven topics. Some topics utilize the same data source. Figure 4 displays six out of seven tabs, and a brief introduction of functionalities to each tab is given as follows.

- Voting History: Latino population and political participation by state (Figure 4 Tab 1). When users click the Voting history tab in the dashboard, the page with title "Latino population and political participation by state (1980-2014)" will appear. The data source for this tab is Reported Voting and Registration by Race and Hispanic Origin, for States. In this page we can facet more than one variables and compare with different ethnic groups and different states over USA. User can select and compare total population, total registered, total citizen, total voted and their percentages as well as their margin of errors. The shadow in the plot indicates margin of errors for percent variables. Users can also download the graph that is of interest.

- Age Group: Ratio of Hispanic over Latino population in Nebraska by age and gender (Figure 4 Tab 2). The data source for this topic is Nebraska Hispanic Population by Sex and Age. In this tab users can explore the pattern of Nebraska Latino population age structure during 1990 to 2013. Users can drag slider button to see a particular year, or click the arrow to display an animation of the line graph. The line graph works similar as a population pyramid. When

Figure 4: Shiny Dashboard: A Demographic Portrait of Latino Population Growth and Past and Potential Political Engagement. The first six tabs are displayed. The dashboard is published online at `https://latino.shinyapps.io/Dashboard/`.

the year changes, a comparison of population age structures between every ten years will also change to the year selected. A static time series plot is also added to display the Hispanic and Non-hispanic adult proportions by sex from 1990 to 2013.

- County of NE population: Proportion of Latino people in Nebraska counties (Figure 4 Tab 3). The data source for this topic is Nebraska Hispanic Population by county. **ggvis** is the engine for the interactions. In this tab users can explore the county level Hispanic population growth in Nebraska. When the user mouses over on the map, there will be a label on the hooked county displayed the total population, total Hispanic and percent Hispanic. If a user clicks a county, then the time series plot of Hispanic population will be updated for that county, and the county will be highlighted in the dot plot, which sorts all the 93 Nebraska

counties by Hispanic population. The county map can be downloaded.

- Greater Omaha Area: Hispanic Citizen Voting Age Population (CVAP) in Nebraska (Figure 4 Tab 4). This tab is made by **leaflet** and uses the data from Citizen Voting Age Group and Cartographic Boundary Shape files. The block group level areal estimated citizen and adult citizen (Potential voter) population in Nebraska were displayed on top of the OpenStreetMap in a Stamen toner-lite style. In this tab users can explore the voting eligibility for Hispanics and other races in all places of Nebraska by zooming in and zooming out.

- Potential Voters: Percentage of Hispanics in All Potential Voters (Adult Citizens) in USA (Figure 4 Tab 5). This map was generated by **plotly** using the Citizen Voting Age Group data. Here users can zoom in and zoom out all the states and compare Hispanic total population, Hispanic adult rate, citizen rate and potential voter rate across USA. The US map can be downloaded as a png file.

- Voters by Race: Percentage of Potential Voters (Adult Citizens) by Race (Figure 4 Tab 6). This tab also utilizes the Citizen Voting Age Group data and is generated by **plotly**. There are four maps generated for four ethnicity groups according to the potential voter rates. Same color scale is applied to all the four maps, so users can easily compare the potential voter rate between races and states. Again, plots can be downloaded.

- Increasing Voters: Increasing Rate of Eligible Voters by Race (Figure 11). The last tab provides a static barchart for comparing the increasing rates of eligible voters among four ethnic groups. The rates were summarized from Hispanic Citizen Voting Age Population.

4. Exploratory Data Analysis

In this section, we use the Shiny Dashboard to explore Hispanic population data. Findings from the interactive graphics are illustrated as follows.

4.1 Midwest versus east/west coast

In Figure 5 we choose to compare large populated states like California and New York with less populated states like Iowa , Missouri and Nebraska. Here we see the total Hispanic population is growing in California and in those states total population is higher than the Midwest but in those states shows their percentage of voted is lower than the large populated states and some percent data are displayed with margin of error (shadow) area.

4.2 Cluster for the middle-aged Hispanics

In Figure 6, the interactive plot shows a big hump in ages 15-30 risen since 1990. The hump firstly appeared for male, then was also shown in female. The static plot, Percentage of Hispanic adult, shows that the population growth of Hispanic males is higher than females. In this plot user can also see a small bump in 2000 to 2004. It may indicate an unusual Hispanic immigrant wave into Nebraska in those years for some reason. As a result, the decreasing bump is also shown in the Non-hispanic population group.

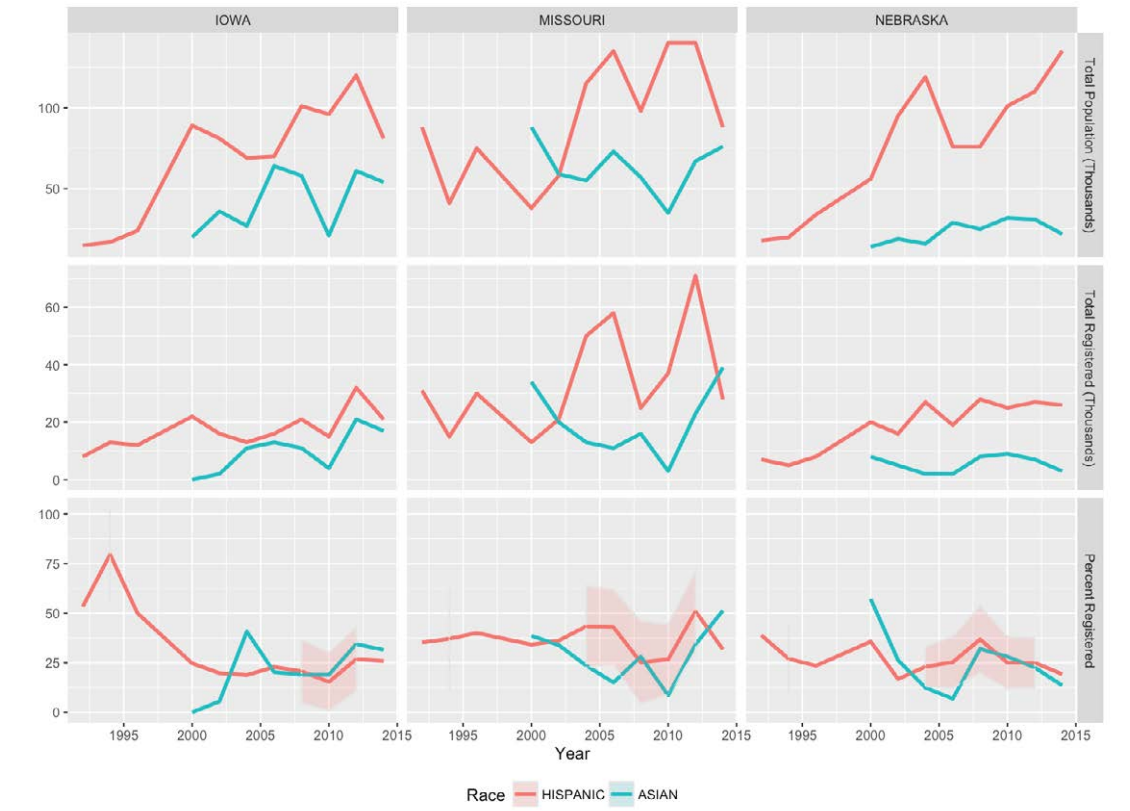

Figure 5: Registration and voting history of Hispanic and other races.

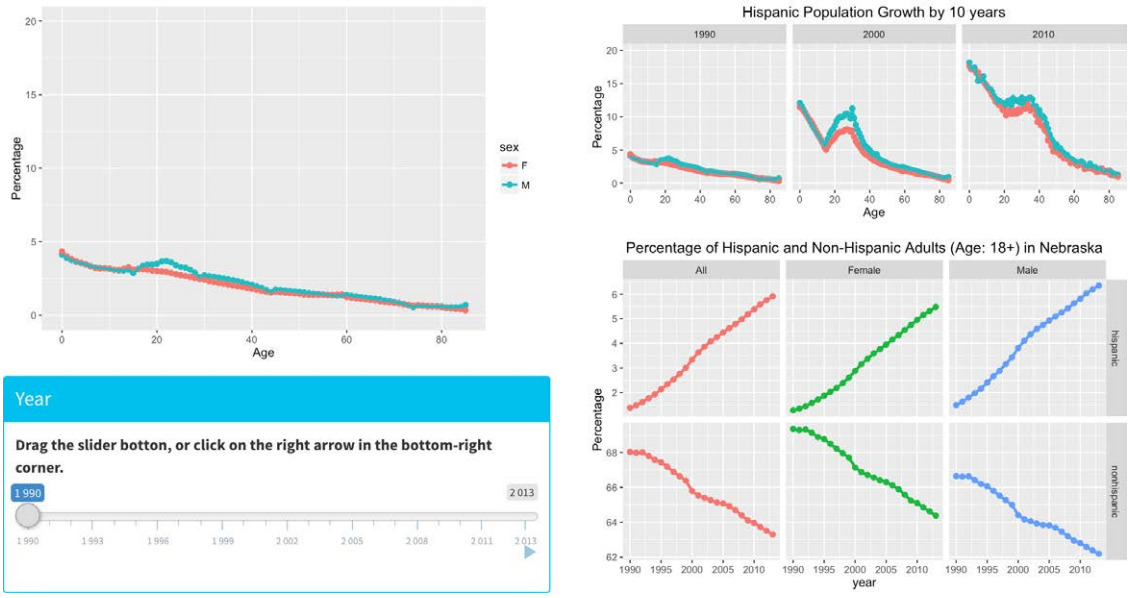

Figure 6: Latino growth rate by sex and age in Nebraska

4.3 Fast growth from the west to the east in Nebraska

Figure 7 shows in 1990 the Hispanic population rate was highest in Scotts Bluff county. At that time the total population of Scotts Bluff was 36025 and Hispanic rate was 14.54% , while the other counties had lower Hispanic rates. If choosing other years, 2000, 2010 or 2013, the map shows that Colfax(43.08%), Dakota (36.55%) and Dawson(32.94%) are the three top most Hispanic populated areas, although in 1990 Hispanic population were less than 7% in those counties. When the year changes, the dark areas in the choropleth map move from the west to the east, which means that the Hispanic population is growing faster in the east counties of Nebraska.

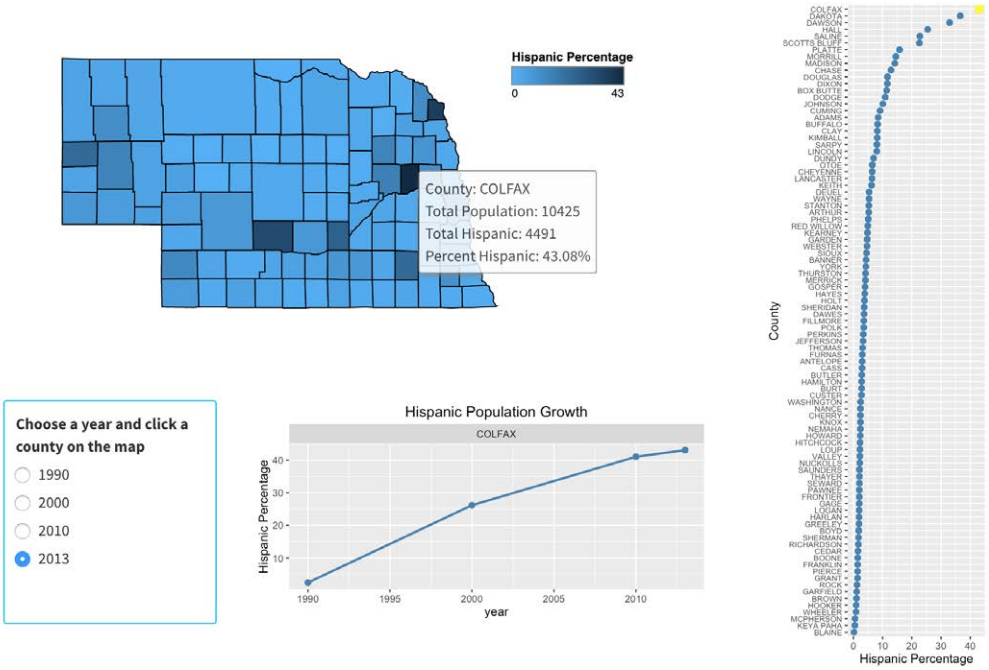

Figure 7: Hispanic population growth in Nebraska counties

4.4 Hispanic populated areas in Nebraska

Figure 8 shows the estimated Hispanic citizen and estimated potential voter rate are higher in Fremont, Ralston , North and East Omaha.

4.5 Large base but low adult citizen rate over US

The interactive map in Figure 9 shows that the adult citizen rate (potential voter rate) of large Hispanic populated state like California is only 42.18% , on the other hand in Vermont total Hispanic is only 9450 (year 2008-2012) but potential voter rate is 60.32% which indicates the highest rate comparing to other states.

4.6 Lowest potential voter rate

Comparing all the graphs we can find that for most states, Hispanic people have the lowest potential voter rate. For example, even for the largest Hispanic populated state, California, the Hispanic potential voter rate was only 42.18%, while the rates for other races were: White 79.81%, African Americans 73.33%, Asian Americans 58.61%, Pacific Islanders 63.96%.

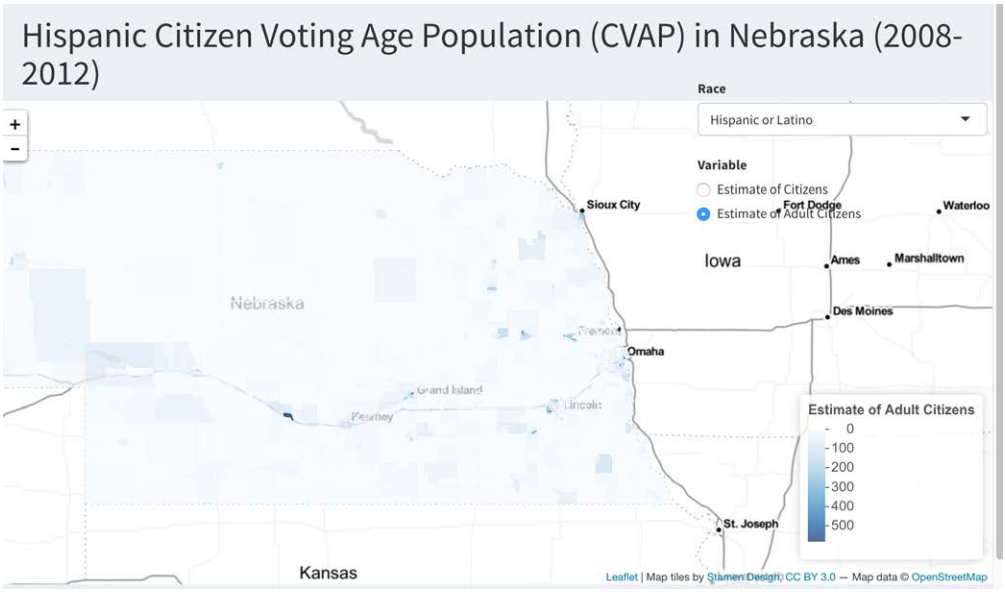

Figure 8: Hispanic population in Nebraska at the block group level

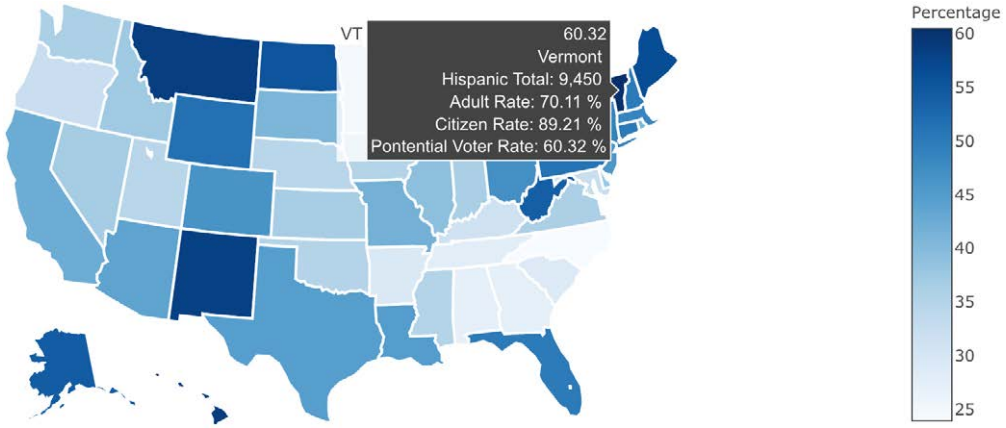

Figure 9: Hispanic Adult Citizen over USA

4.7 Highest increasing rate of eligible voters

Hispanics are the youngest major racial or ethnic group in the United States. About one-third, or 17.9 million, of the national Hispanic population is younger than 18, and about a quarter, or 14.6 million, of all Hispanics are Millennial (ages 18 to 33 in 2014).By comparison, half of the black population and 46% of the U.S. Asian population are Millennial or younger. Among whites, the national oldest racial group, only about four-in-ten are Millennial or younger (39%). For that reason new eligible Hispanic voter rate is higher than the others.

The number of Hispanic eligible voters has grown at one of the fastest clips of any group over the past eight years and is projected to be 40% higher in 2016 than in 2008, according to a new Pew Research Center analysis of government data [5].

Figure 11 shows that in recent year (2012-2016) the growth rate of new eligible voter is (17%) which is highest in Hispanic population group than others. In this sense, the Hispanic voters may

[5] `http://www.pewresearch.org`

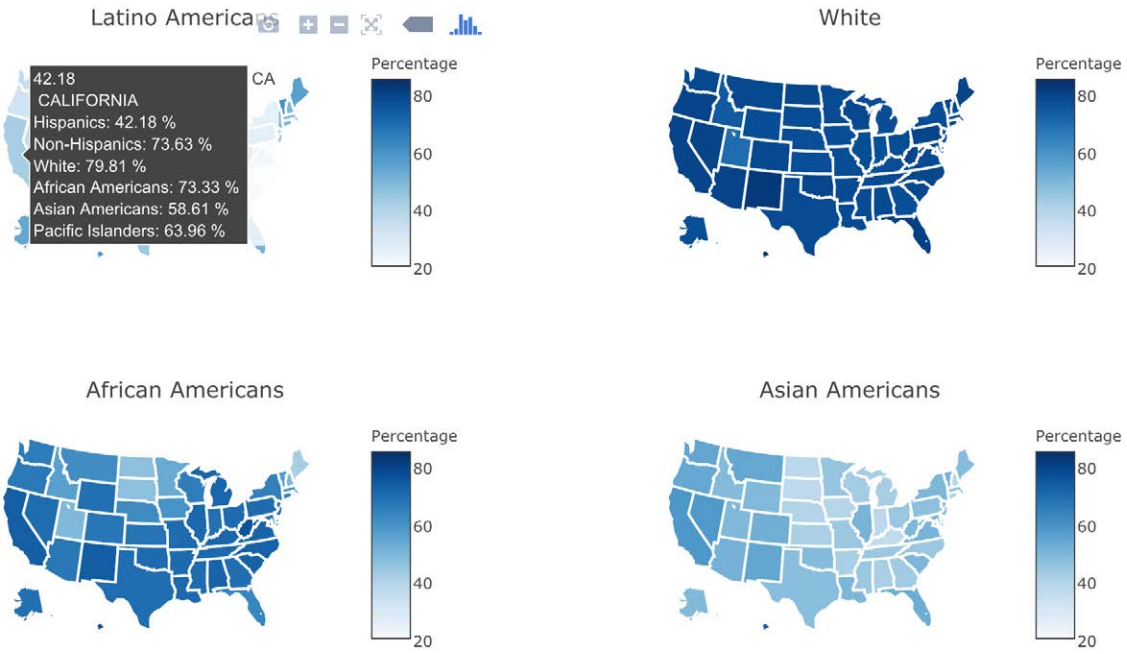

Figure 10: Potential voter rate over USA

be play a vital role in 2016 election.

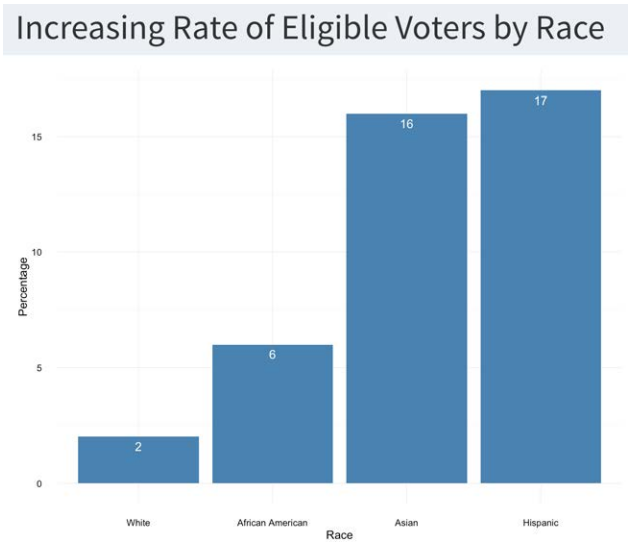

Figure 11: New eligible voter increasing rate

5. Concluding remarks

5.1 Conclusion

This research intended to explore and reveal patterns in the growth of Hispanic population in Nebraska and across USA. Data were collected from the US Census Bureau and OLLAS. All the issues and mistakes in the data were properly solved, and a shiny dashboard was developed.

With the dashboard, users can interact with data from different aspects, compare features among Hispanic and other races, visualize the population growth temporally and geographically, and learn the patterns of Hispanic political participation.

Our work also reveals characteristics of Hispanic population in Nebraska and USA. Below is a summary of our findings: Hispanic population and their voting position rank the third comparing to three other races. There was a big hump in ages 15-30 risen since 1990 for Nebraska Hispanics. Hispanic male population growth rate is higher than female population. In Nebraska, Hispanic population is growing faster in east counties than the west. Hispanic estimated citizen rate and estimated adult citizen rate are higher in Omaha and Fremont than other cities in Nebraska. In some states, although total Hispanic population is higher but there potential voter rate is lower than other races. Hispanic new eligible voter rate is significantly increased from 2012 to 2016, which would be a crucial factor to the 2016 election.

5.2 Future work

In future, we plan to (1) collect data from Iowa to complete the research for Omaha metropolitan area; (2) develop more functionality to the dashboard for flexibly reading and saving data; (3) add time series analysis for the historical data and predict the pattern in political participation; and (4) extend the work to other ethnic groups or other census data.

Acknowledgments

This research was partially supported by the Office of Latino/Latin American Studies (OLLAS) in University of Nebraska at Omaha. We thank Dr. Juan Cases, Dr. Lissette Aliaga-Linares, Dr. Jasney Cogua-Lopez, Dr. Mahbubul Majumder, and Dr. John Konvalina for their critical comments and substantial suggestions. We also thank the authors of **R** packages **shiny**, **shinydashboard**, **reshape**, **plyr**, **ggplot2**, **ggvis**, **plotly**, and **leaflet**.

REFERENCES

Aliaga Linares, Lissette and Jasney Cogua-Lopez (2016), "Latinos and the Economic Downturn in Nebraska: Demographic and Socioeconomic Trends 2005-2013/2014," Office of Latino/Latin American Studies (OLLAS) at the University of Nebraska at Omaha, Omaha, NE

Cogua-Lopez, Jasney, Lissette Aliaga-Linares and Lourdes Gouveia (2015), "Latinos throughout the City: A Snapshot of Socio-demographic Differences in Omaha, Nebraska," Office of Latino/Latin American Studies (OLLAS) at the University of Nebraska at Omaha, Omaha, NE

R Core Team (2016), "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, `http://www.R-project.org/`

Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2016), "shiny: Web Application Framework for R," R package version 0.14, `https://CRAN.R-project.org/package=shiny`

Winston Chang (2016), "shinydashboard: Create Dashboards with 'Shiny'," R package version 0.5.2, `https://CRAN.R-project.org/package=shinydashboard`

Hadley Wickham (2007), "Reshaping Data with the reshape Package," *Journal of Statistical Software*, 21(12), 1-20, `http://www.jstatsoft.org/v21/i12/`

Hadley Wickham (2011), "The Split-Apply-Combine Strategy for Data Analysis," *Journal of Statistical Software*, 40(1), 1-29, `http://www.jstatsoft.org/v40/i01/`

Hadley Wickham (2009). *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.

Winston Chang and Hadley Wickham (2016), "ggvis: Interactive Grammar of Graphics," R package version 0.4.3, `https://CRAN.R-project.org/package=ggvis`

Carson Sievert, Chris Parmer, Toby Hocking, Scott Chamberlain, Karthik Ram, Marianne Corvellec and Pedro Despouy (2016), "plotly: Create Interactive Web Graphics via 'plotly.js'," R package version 3.6.0, `https://CRAN.R-project.org/package=plotly`

Joe Cheng and Yihui Xie (2016), "leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library," R package version 1.0.1, https://CRAN.R-project.org/package=leaflet