MGTA602 – Data Visualization Fall 2022

Assignment 1 – Segmenting Clinton and Obama Voters

Sarah Mansoor

225825060

October 29, 2022

Table of Contents

Executive Summary

The elections in the United States are underway in 2008. Barack Obama is taking the lead the week of February 12, 2008, in the delegate count. The winner of the Democratic Party's presidential primaries will face the Republican Party's nominee in the election to become the next president of the United States. Obama's primary opponent is Hillary Clinton who is running ads in Ohio. Clinton's "Night Shift" speech and Obama's "Down on the Farm" speech caused controversy on whether they knew the demographic they were targeting.

The Problem

As of February 19, 2018, 1,131 counties had yet to report their county-level votes. Clinton and Obama's next strategies would have an impact on how they could influence voters. Which of these factors though was the question. Voters are represented by several factors: gender, age, ethnicity, race, education, income, employment status, languages spoken, government dependency, disabilities, home ownership status, and mobility. I will determine which factors impact the voters' decision so that Clinton and Obama can strategize for their upcoming campaigns.

The data

The data is provided by the University of Virginia and consists of 41 columns and 2868 rows. The data available is sufficient to analyse the question at hand, however it might be helpful to have the votes of any other Democratic candidates in counties as well. But this should not affect analysing how impactful Obama and Clinton's strategies are.

Analysis and Results

To begin my analysis, I looked at the number of votes for Clinton and Obama to see the distribution of spread. From figure 1, Clinton has 10.3 million votes and Obama has 10.7 million votes. There has been a total of 22.3 million votes, and Clinton and Obama make up 21.1 million of these. This suggests that 0.2 million votes are being cast to other democratic leaders. Next, I am interested in seeing the distribution

of their votes by state. California has the highest number of votes for both Clinton and Obama, suggesting that California is a democratic state so Clinton and Obama do not necessarily need to run as many campaigns there since they will get democratic votes anyways. Clinton had more votes in California than Obama (see Figure 2). A state that Clinton should pay attention to is Illinois because Obama currently has 1.3 million votes there while Clinton only has 0.6 million. Moving on to their votes by region, Clinton needs to run more ads in the Midwest since she only has 1.1 million votes compared to Obamas 1.9 million and likewise Obama needs to run more ads in the Northeast and West as he has an average of 0.4 million votes less in those regions (see Figure 3). Now that I understood what the distribution of their votes looked like geographically, I want to see what kinds of demographics each should focus their campaigns on. Figure 4 looks at the distribution of the rate of bachelor's degree holding folks and who they voted for and from this I gather that counties with an average of 23% holding a bachelor's degree will most likely vote for either Clinton or Obama. Counties with higher rates have higher vote counts for Obama, this suggests that Obama should continue campaigning in counties with higher bachelor's degree rates to increase votes, but Clinton should even more so. Now that I have looked at a bachelor's degree, what about a high school diploma? Sure enough, the average rate for counties with high school diplomas is much higher than a bachelor's degree. Counties with an average of 82% with a high school degree see the most votes for Clinton and Obama. Again, these votes go down as the rate of high school degrees goes up for Clinton, thus she needs to target more counties with higher education rates all in all (see figure 5). Very few votes of Obama's and Clintons' come from the disabled, they both need to implement better solutions for the disabled to aid and help them and in turn gain their votes (see Figure 6). Most votes were casted by males as the number of males per 100 females increased in counties, Obama gained more voters. Clinton should focus on counties with more males than females to encourage females to vote (see Figure 7). Figures 8 and 9 look at the distribution of votes related to income. From figure 8, the highest votes for both come from counties with an average income of 34K. Both Clinton and Obama need to shift their campaign

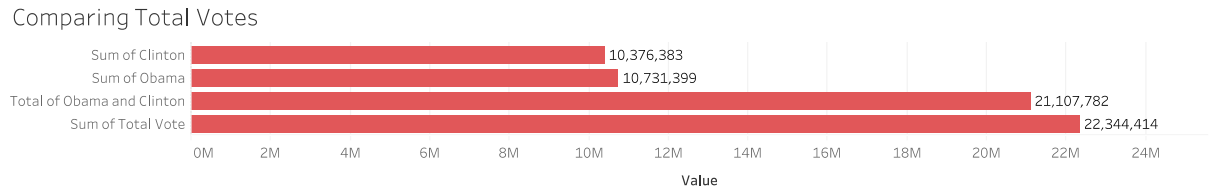
strategies towards lower incomes as they both have very few votes in that demographic. Figure 9 illustrates that counties with a lower rate of income above 75K get more votes for Clinton, so Obama should target that demographic more. Lastly, figures 10, 11 and 12 look at the distribution of Asian, Black, and White voters. From figure 10, most of Clinton and Obamas votes are from non-Asians as the mode rate is 0. Both Clinton and Obama need to adjust their campaigns to benefit Asian-American communities. From figure 11, the mode rate for Black voters is 3 for Obama and Clinton. Both their votes significantly decrease as the rate of black people gets higher in counties; they both need to demonstrate activism for the black community in America to securely gain their votes. Obama still has more votes from the African American community as the rate increases, but it is not much of improvement from Clinton. The mean rate of White voters for both Clinton and Obama is around 90 which is much higher than any of the Black and Asian votes (see Figure 12). This suggests that Clinton and Obama shift their campaigns to counties that have lower rates of White voters to reach more demographics.

Conclusion

Both democratic leaders need to change their campaign strategies if they want to win. Clinton needs to focus on Illinois, the Midwest region, counties with higher bachelor's degree and high school diploma rates, the disabled, counties with more males than females, lower income communities, and the Black and Asian community. Obama needs to shift his focus to the Northeast and West, the disabled, lower incomes, and Black and Asian communities. Clinton has a lot more work to do than Obama, giving him a higher chance of winning, which is exactly that happened.

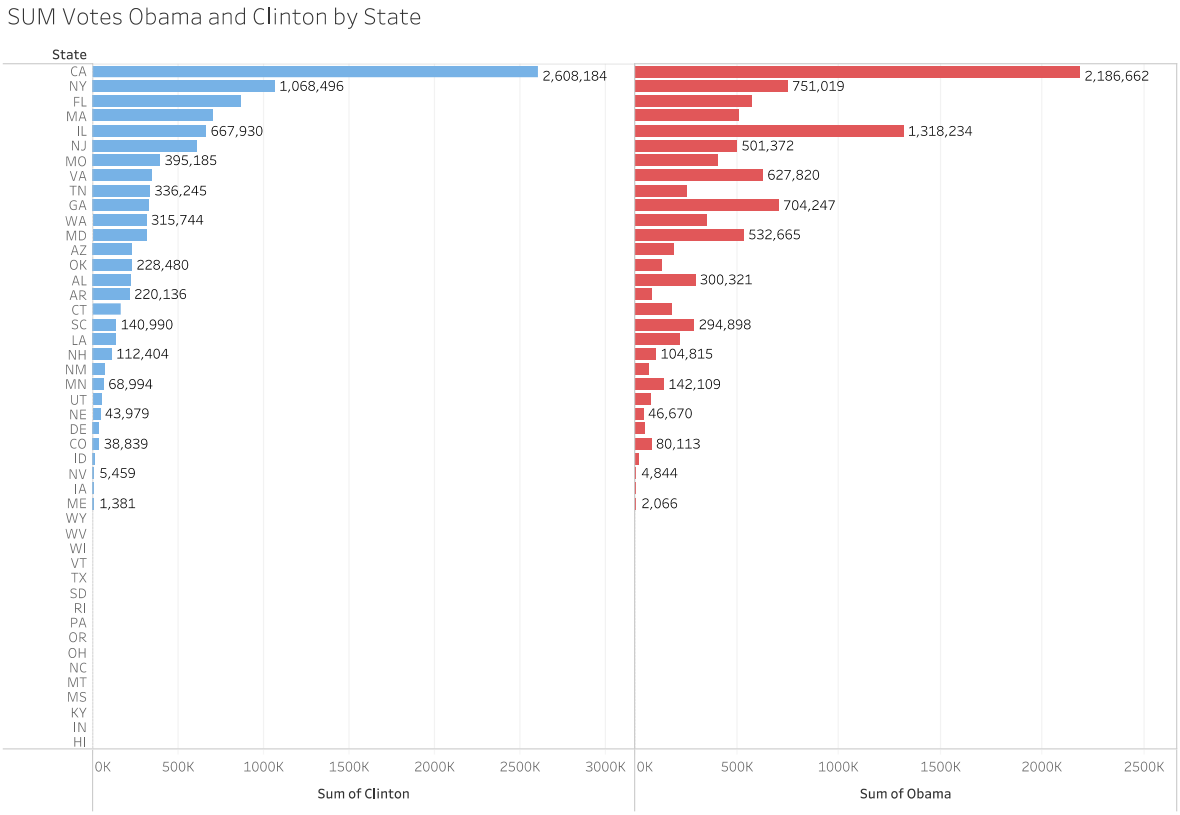
Appendix

Figure 1



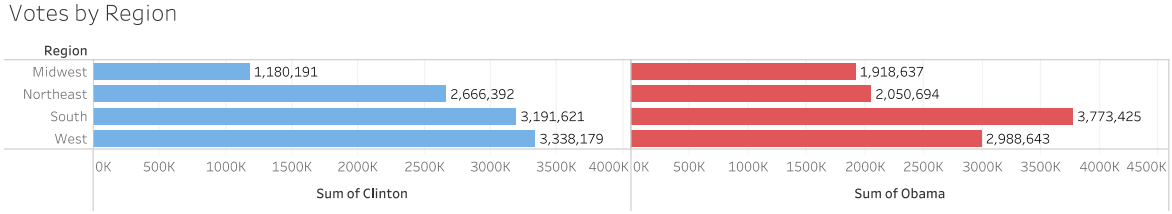
Sum of Clinton, sum of Obama, Total of Obama and Clinton and sum of Total Vote.

Figure 2



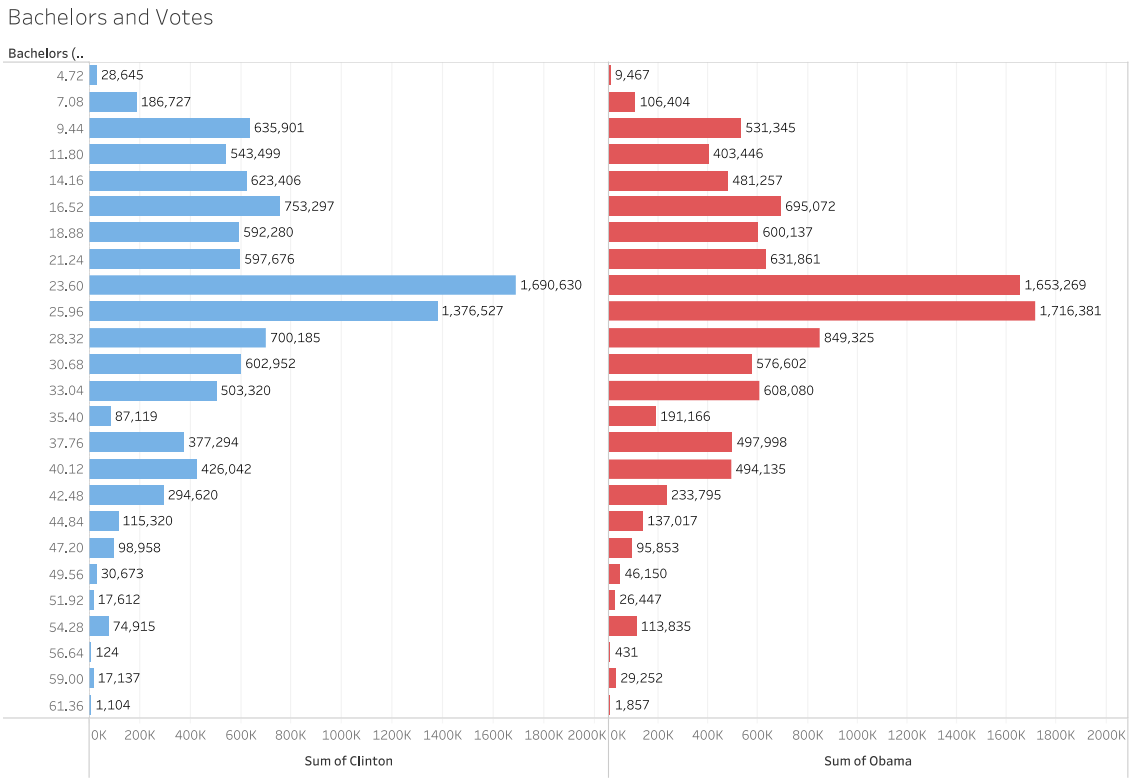
Sum of Clinton and sum of Obama for each State. For pane Sum of Clinton: The marks are labeled by sum of Clinton. For pane Sum of Obama: The marks are labeled by sum of Obama.

Figure 3



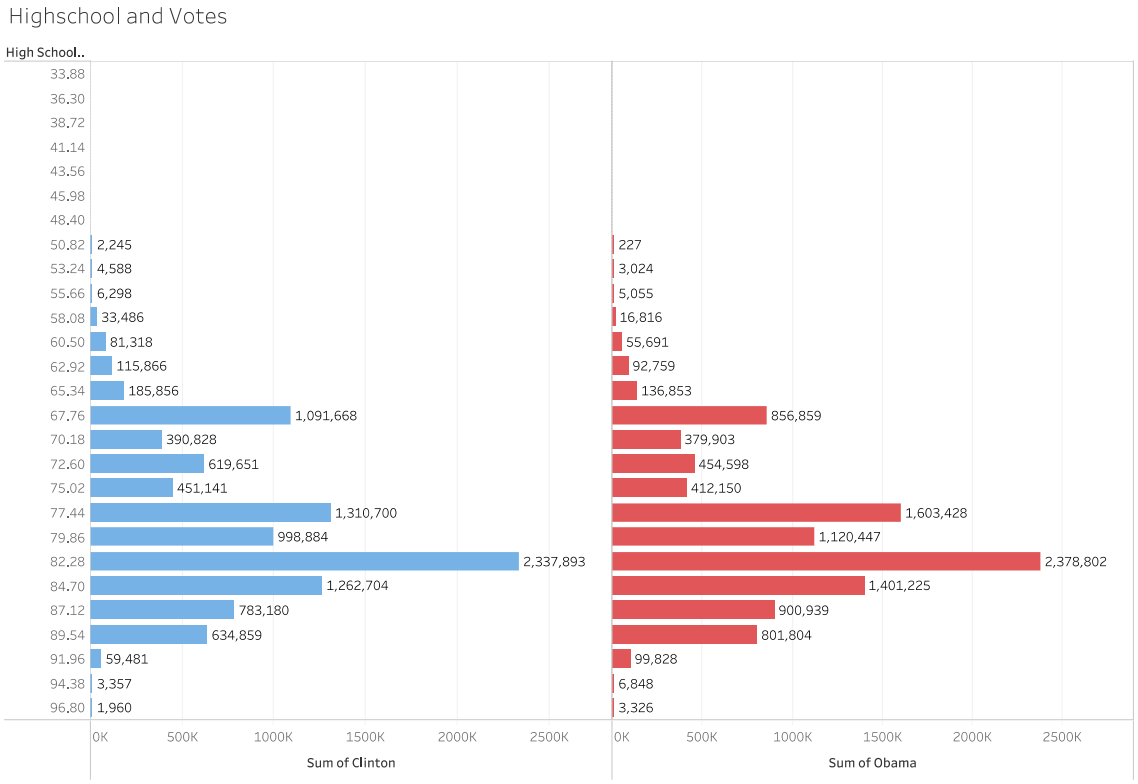
Sum of Clinton and sum of Obama for each Region. For pane Sum of Clinton: The marks are labeled by sum of Clinton. For pane Sum of Obama: The marks are labeled by sum of Obama.

Figure 4



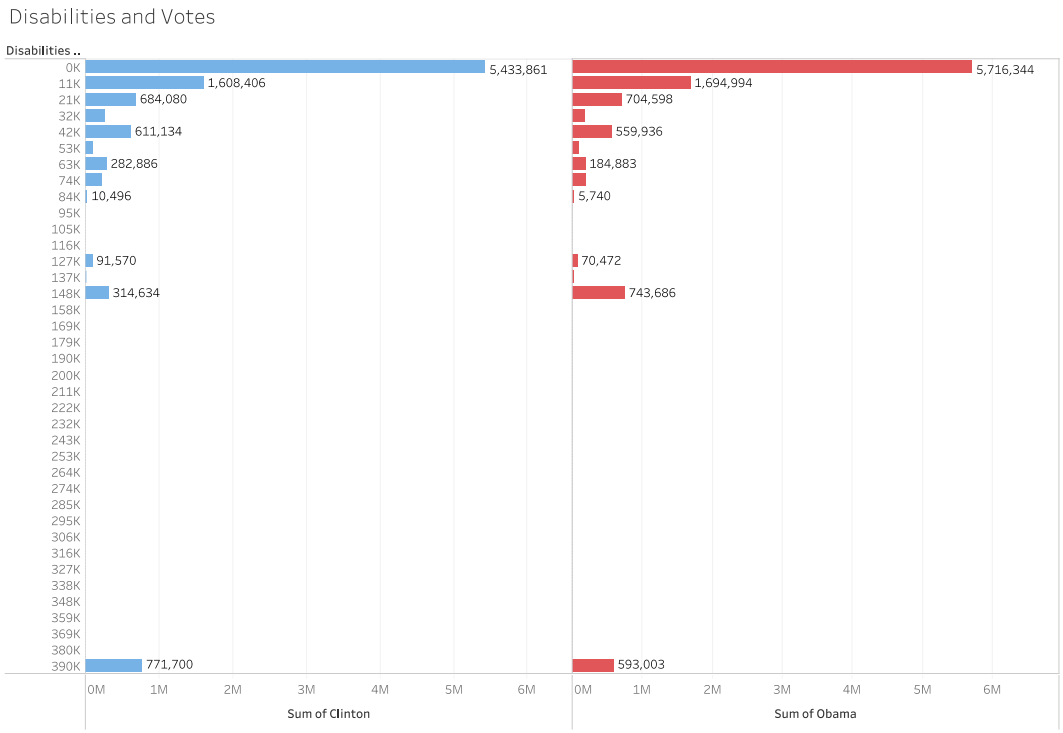
Sum of Clinton and sum of Obama for each Bachelors (bin). For pane Sum of Clinton: The marks are labeled by sum of Clinton. For pane Sum of Obama: The marks are labeled by sum of Obama. The view is filtered on Bachelors (bin), which excludes Null.

Figure 5



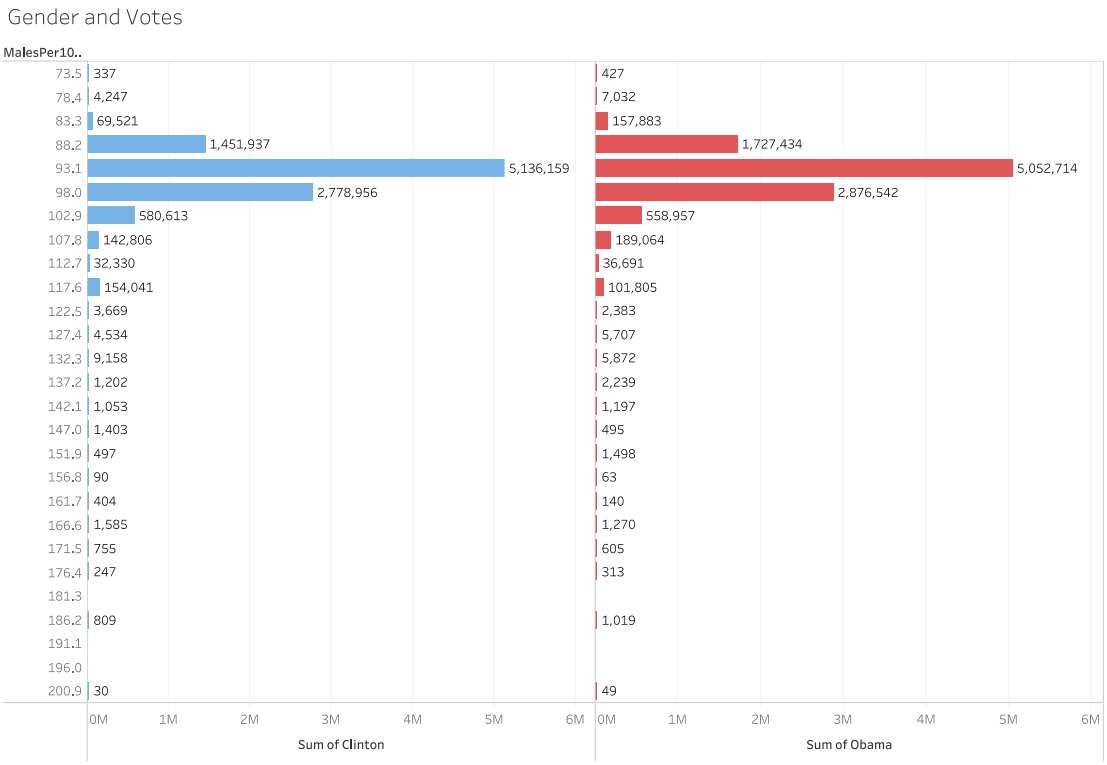
Sum of Clinton and sum of Obama for each High School (bin). For pane Sum of Clinton: The marks are labeled by sum of Clinton. For pane Sum of Obama: The marks are labeled by sum of Obama. The data is filtered on Bachelors (bin), which excludes Null.

Figure 6



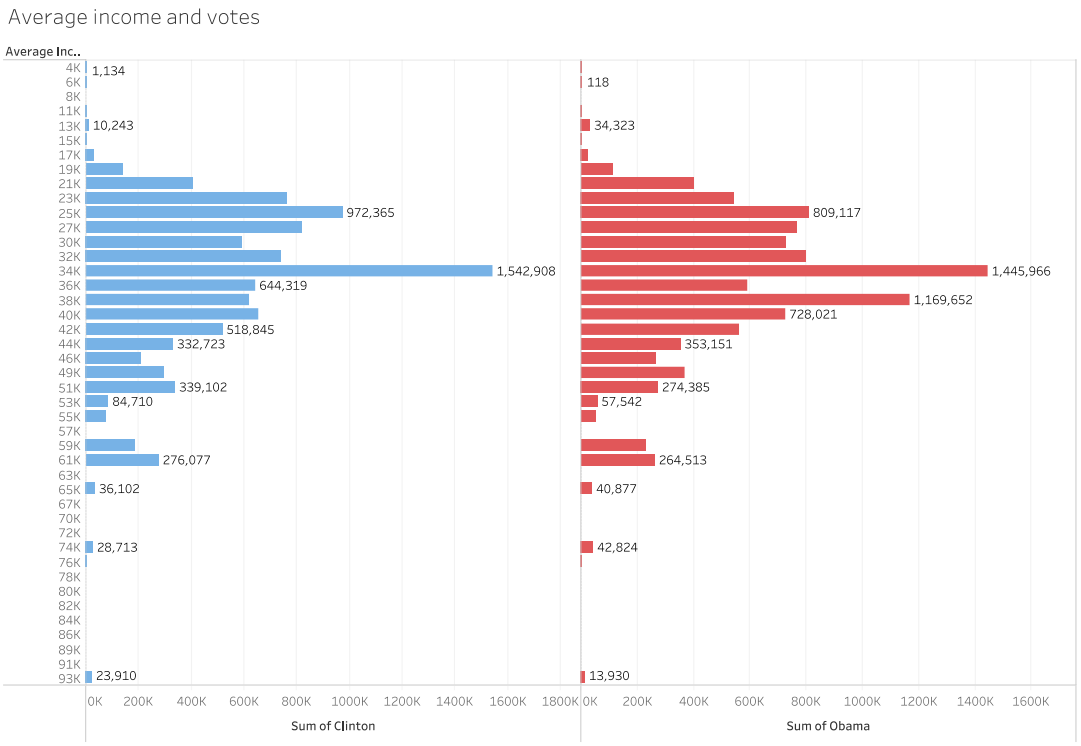
Sum of Clinton and sum of Obama for each Disabilities (bin). For pane Sum of Obama: The marks are labeled by sum of Obama. For pane Sum of Clinton: The marks are labeled by sum of Clinton. The view is filtered on Disabilities (bin), which excludes Null.

Figure 7



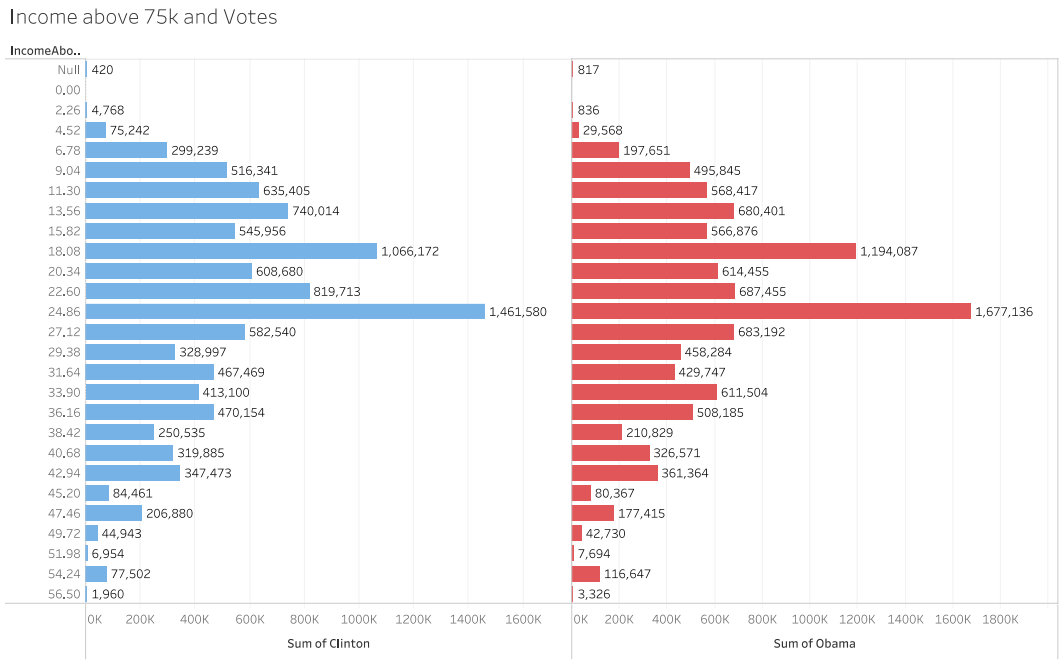
Sum of Clinton and sum of Obama for each MalesPer100Females (bin). For pane Sum of Obama: The marks are labeled by sum of Obama. For pane Sum of Clinton: The marks are labeled by sum of Clinton.

Figure 8



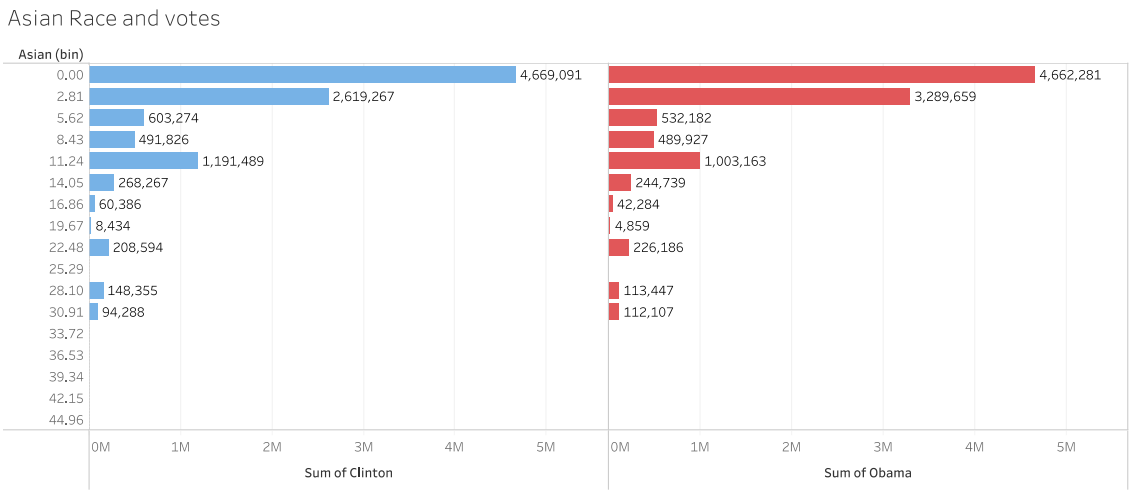
Sum of Clinton and sum of Obama for each Average Income (bin). For pane Sum of Obama: The marks are labeled by sum of Obama. For pane Sum of Clinton: The marks are labeled by sum of Clinton. The view is filtered on Average Income (bin), which excludes Null.

Figure 9



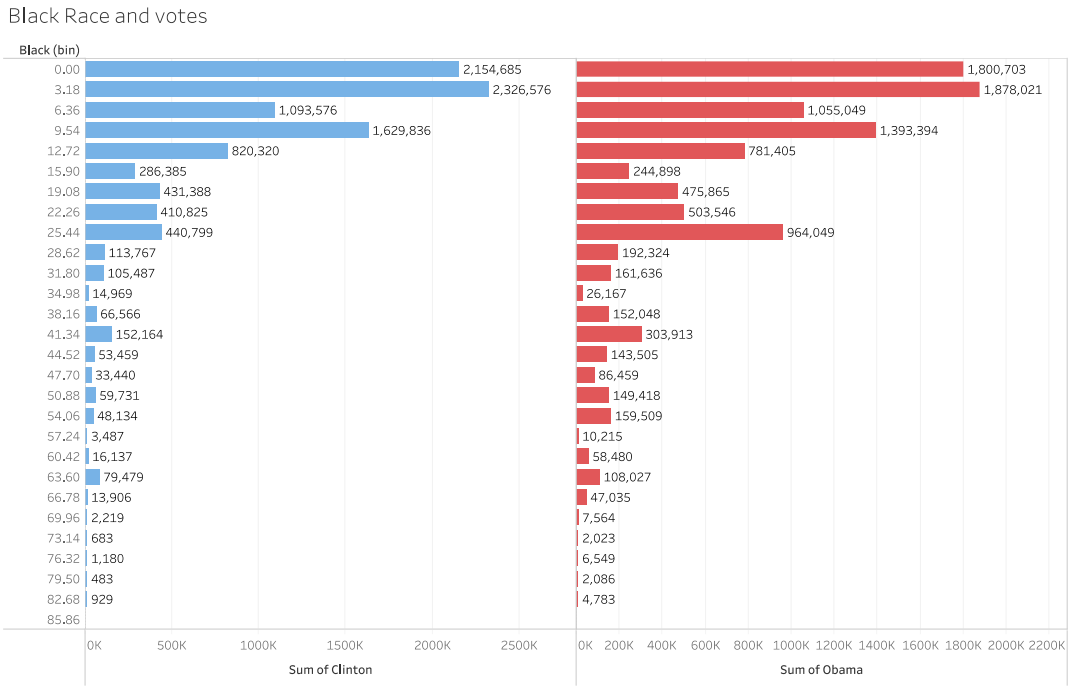
Sum of Clinton and sum of Obama for each IncomeAbove75K (bin). For pane Sum of Obama: The marks are labeled by sum of Obama. For pane Sum of Clinton: The marks are labeled by sum of Clinton.

Figure 10



Sum of Clinton and sum of Obama for each Asian (bin). For pane Sum of Obama: The marks are labeled by sum of Obama. For pane Sum of Clinton: The marks are labeled by sum of Clinton. The view is filtered on Asian (bin), which excludes Null.

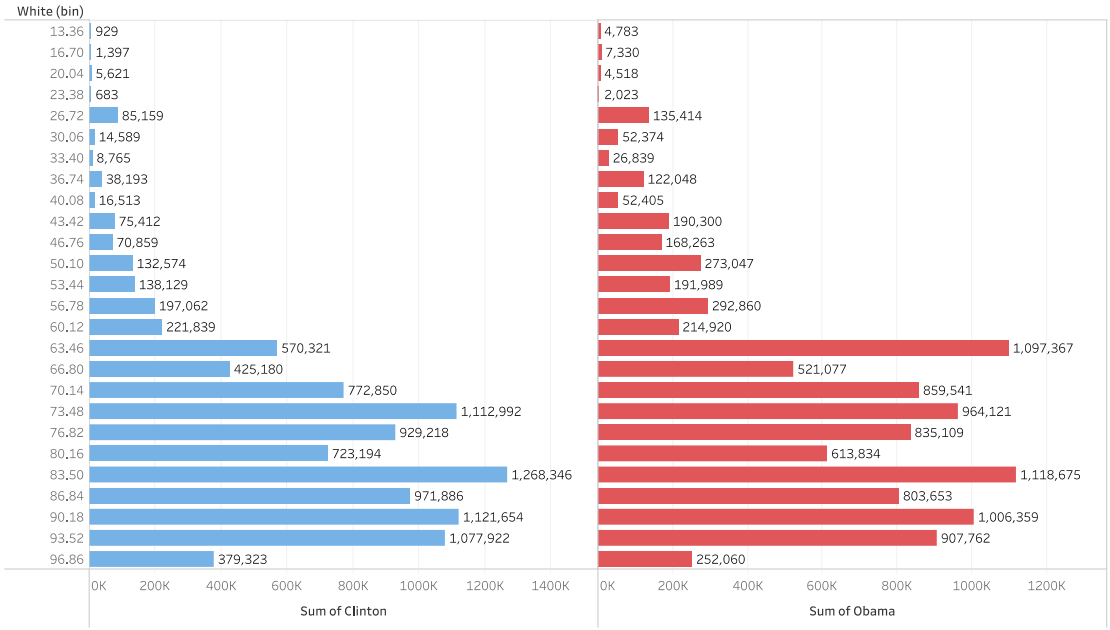
Figure 11



Sum of Clinton and sum of Obama for each Black (bin). For pane Sum of Obama: The marks are labeled by sum of Obama. For pane Sum of Clinton: The marks are labeled by sum of Clinton. The data is filtered on Asian (bin), which excludes Null. The view is filtered on Black (bin), which excludes Null.

Figure 12

White Race and votes



Sum of Clinton and sum of Obama for each White (bin). For pane Sum of Obama: The marks are labeled by sum of Obama. For pane Sum of Clinton: The marks are labeled by sum of Clinton. The data is filtered on Asian (bin) and Black (bin). The Asian (bin) filter excludes Null. The Black (bin) filter excludes Null.