Exploratory Data Analysis of Election and Socioeconomic County Information in 2020

Ainsley Carlson

2025-09-24

Purpose

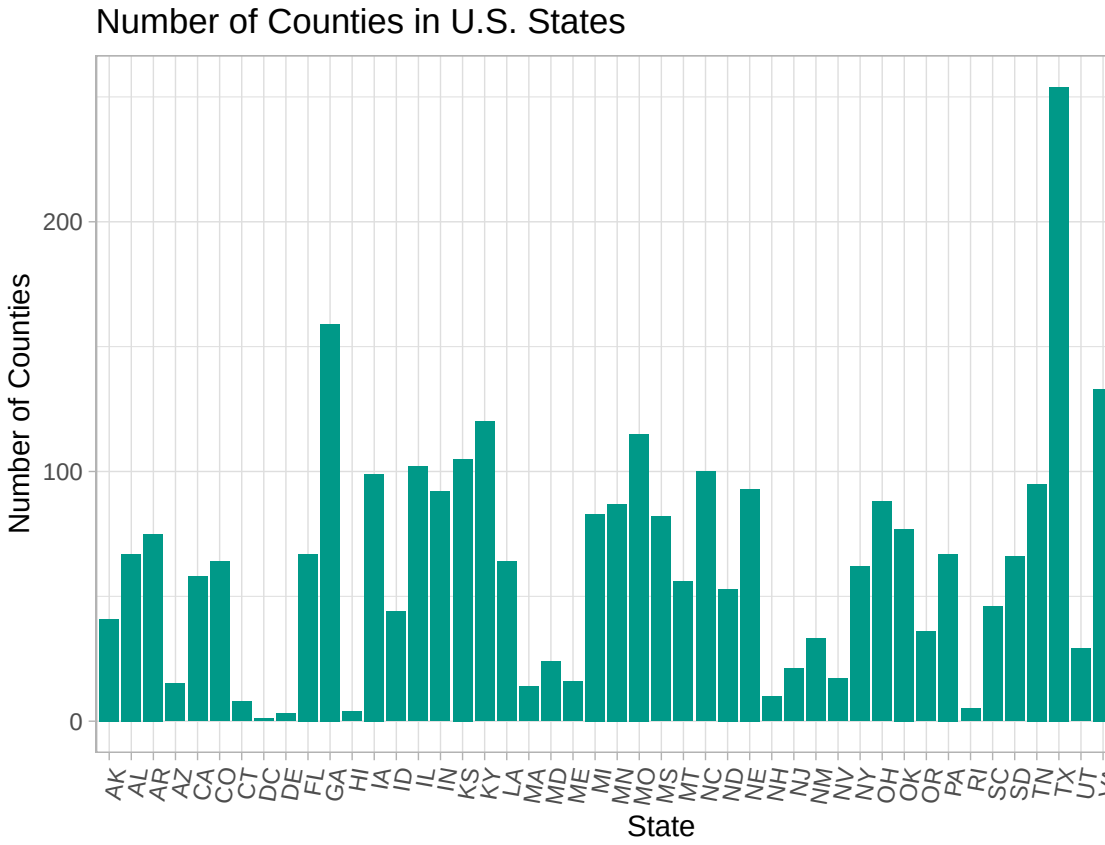
The purpose of this is to explore the relationship of how different socioeconomic factors in United States (U.S.) counties vary in 2020 based on the winning presidential party in that county. The socioeconomic variables of interest are the unemployment rate, poverty rate educational attainment among counties. Additionally, information about the counties is also included to understand more about the data. No visualizations were made for county names as some county names appear in multiple states, and there are also no visualizations for FIPS codes as they are all uniuque.

Data Information

The data source was generated by combining data of United States presidential elections, unemployment, poverty, and educational attainment. Data was combined on the county level and filtered to the years 2019 and 2020. Below is information about the dataset including size and variable descriptions Size: 3269 Observation of 12 variables

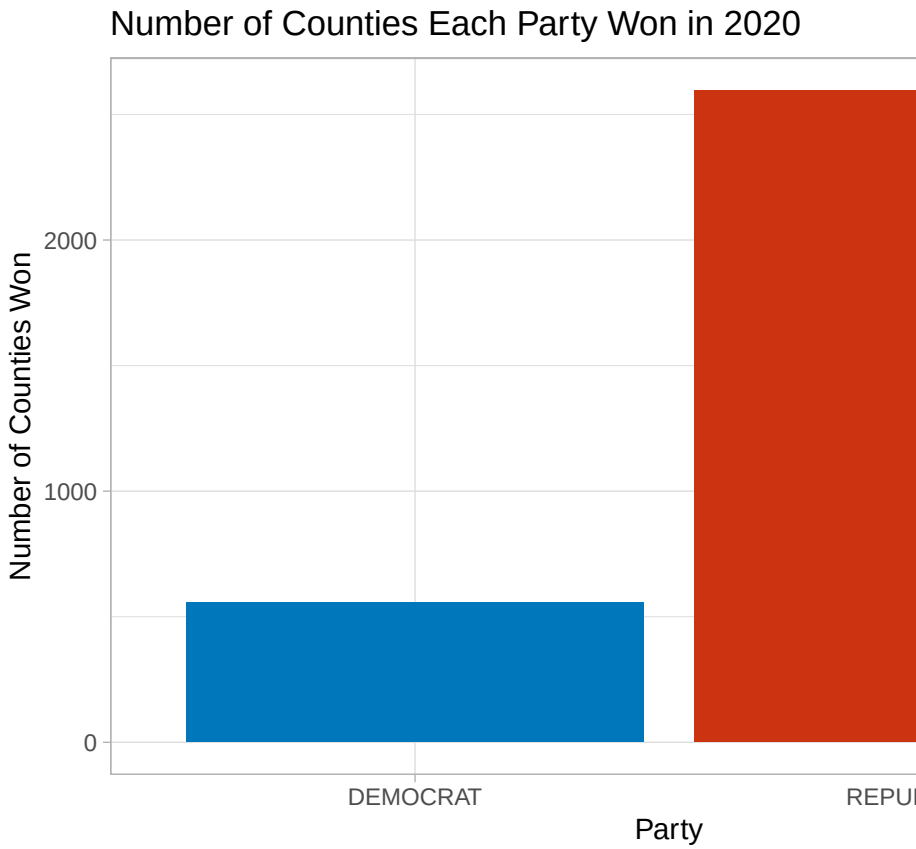
Variable Description State: [Character] State name
state_po: [Character] State Abbreviation
county_name: [Character] Name of County Fips_Code: [Numeric] County FIPS Code
party: [Character] Winning party of the 2020 presidential election in the county
partyvotes: [Numeric] Number of votes towards the winning party in the county
unemployment_rate: [Numeric] Unemployment rate of county in 2020
poverty_rate: [Numeric] Poverty rate of county in 2019 less_highschool: [Numeric] Percentage of people in the county whose highest level of education is less than a high school diploma
highschool: [Numeric] Percentage of people in the county whose highest level of education is a high school diploma
some_college: [Numeric] Percentage of people in the county whose highest level of education is some college/ an associate's degree bachelor_more: [Numeric] Percentage of people in the county whose highest level of education is less than a college degree or higher State: tells us how many counties in each state state_po: tells us how many counties in each state county_name: not going to give much info[Fips_Code: all unique values
party: how many counties that party won partyvotes: number of total for winning party unemployment_rate: histogram of unemployment rate distribution poverty_rate: histogram of povrty rate distribution less_highschool: [Numeric] Percentage of people in the county whose highest level of education is less than a high school diploma
highschool: [Numeric] Percentage of people in the county whose highest level of education is a high school diploma
some_college: [Numeric] Percentage of people in the county whose highest level of education is some college/ an associate's degree bachelor_more: [Numeric] Percentage of people in the county whose highest level of education is less than a college degree or higher

County Information

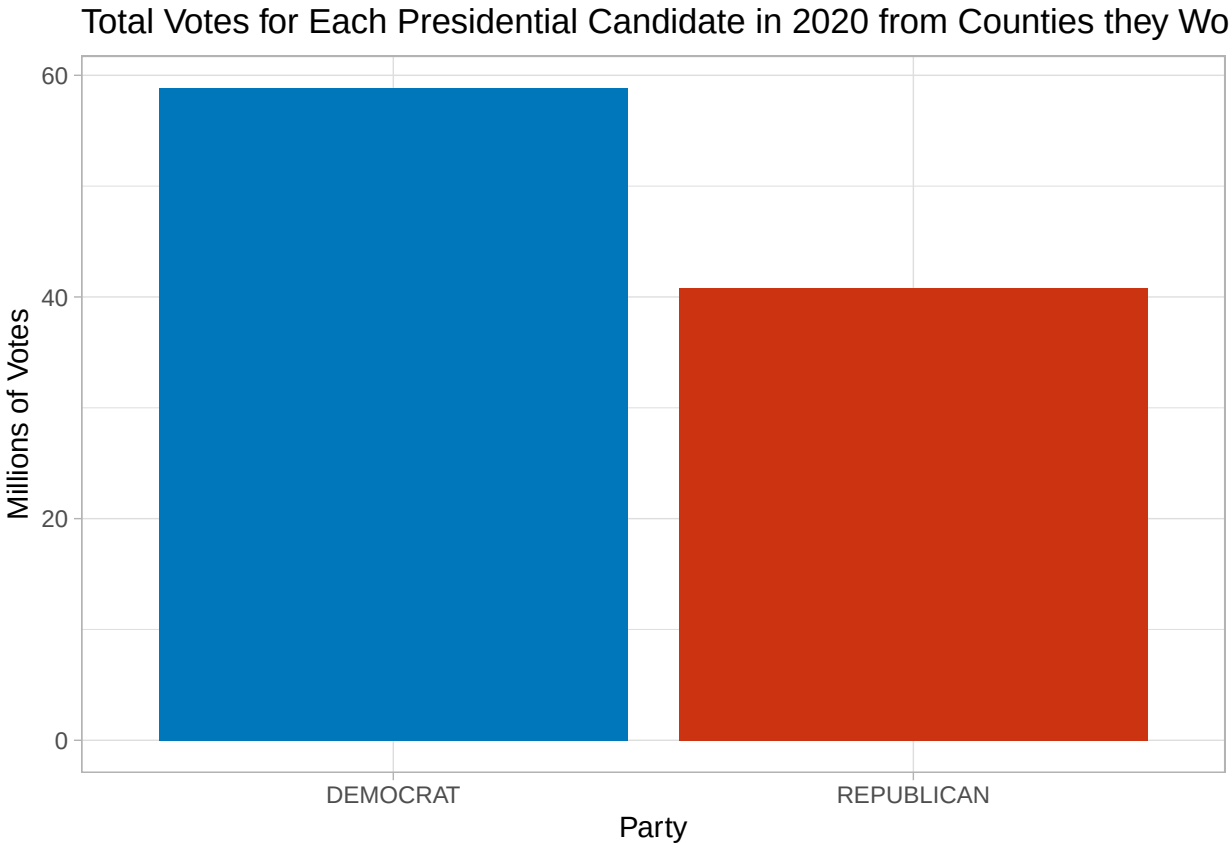


Counties in U.S. States

Looking at the number of counties in each United States shows that Texas has the most counties with over 250, followed by Florida with more than a little more than 150 and Virginia with a little less than 150. Excluding D.C., Delaware, Rhode Island, Connecticut and Hawaii appear to have the least amount of counties with about 10 or less. It is important to remember that the number of counties a state has is not related to how many people live in that state.

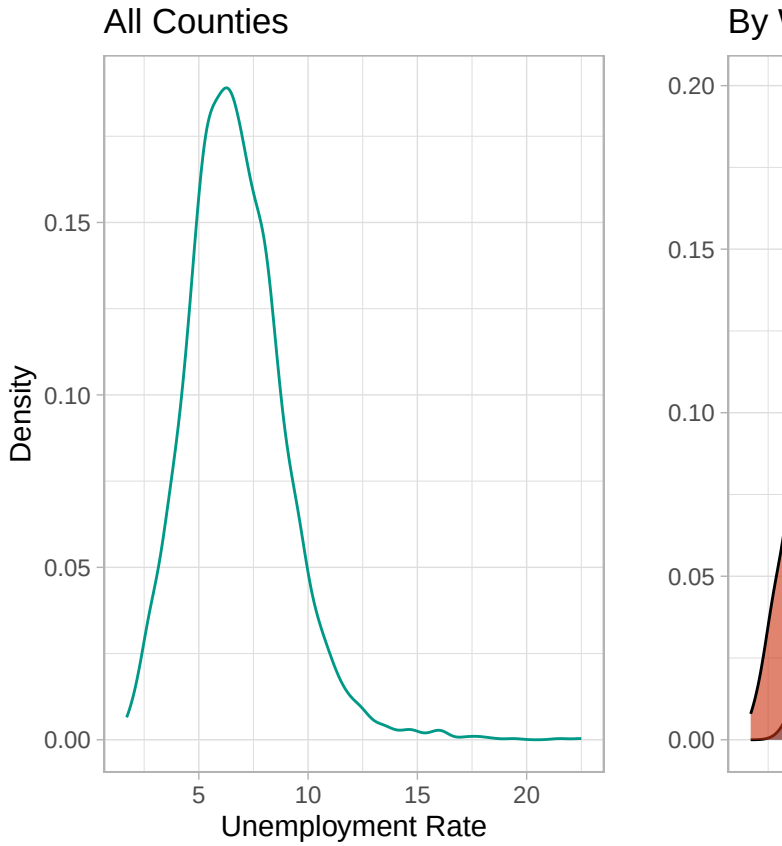


Winning Parties by Counts and Votes



The top bar graphs above emphasize the point about the number of counties not corresponding to the population size. The first bar chart shows that in 2020 over 2500 counties voted for the republican presidential candidate compared to a little over 500 counties voting for the democratic presidential candidate. However, in 2020 the democratic candidate was the winning candidate in 2020 in both the popular vote and electoral vote. This is seen looking at the votes total votes each presidential candidate had in the 2020 election only from counties they won shown in the bottom bar graph above. The bar graph shows that the democratic candidate has more of the total votes as seen in the graph below. This does not tell the whole story as it does not include all the votes, but it shows why it is important just to look at the number of counties the candidate won to determine who won the election.

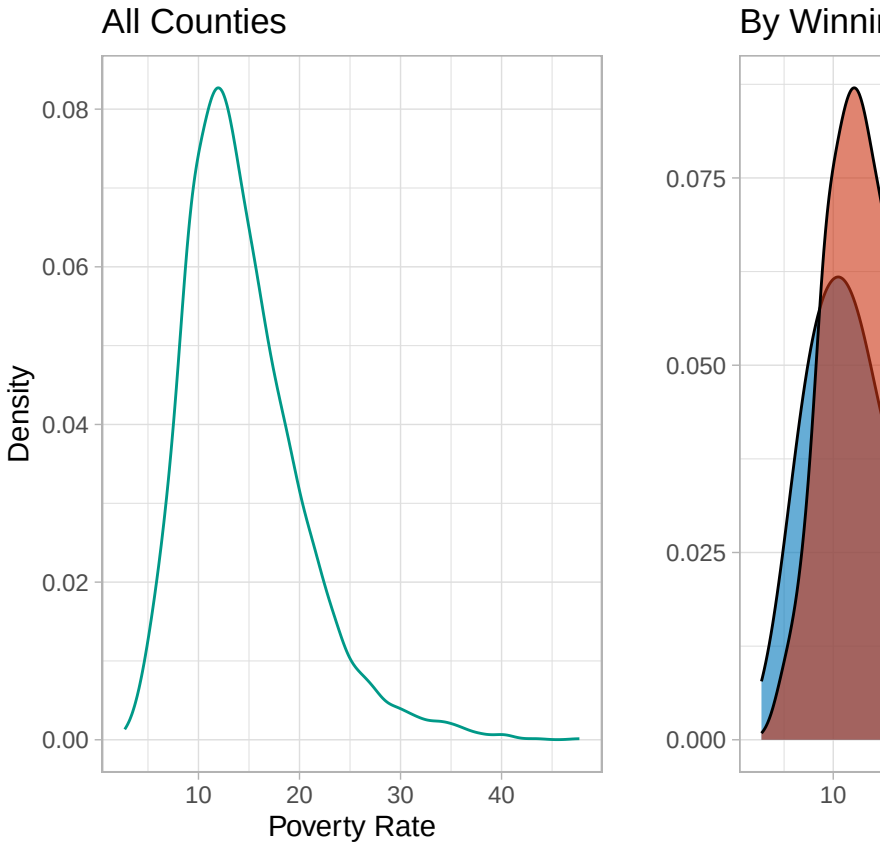
Unemployment Rates



Density Curves of Unemployment Rates in 2020

Above, the density curve of the unemployment rate shows a peak around 7% unemployment rate for about 18-19% of U.S. Counties. Most U.S. counties appear to have an unemployment rate less than 10% with a few counties having an unemployment rate greater than 15%, which may have been due to the Covid-19 pandemic. Separating the unemployment rate by the winning party of the county shows that democratic voting counties had a slightly higher peak unemployment rate than republican voting counties. Around 20% of democratic voting counties had an unemployment rate of 8%, while around 20% of republican voting counties had an unemployment rate of 6%. Most of the counties with an unemployment rate greater than 15% appear to be those that are democratic voting counties.

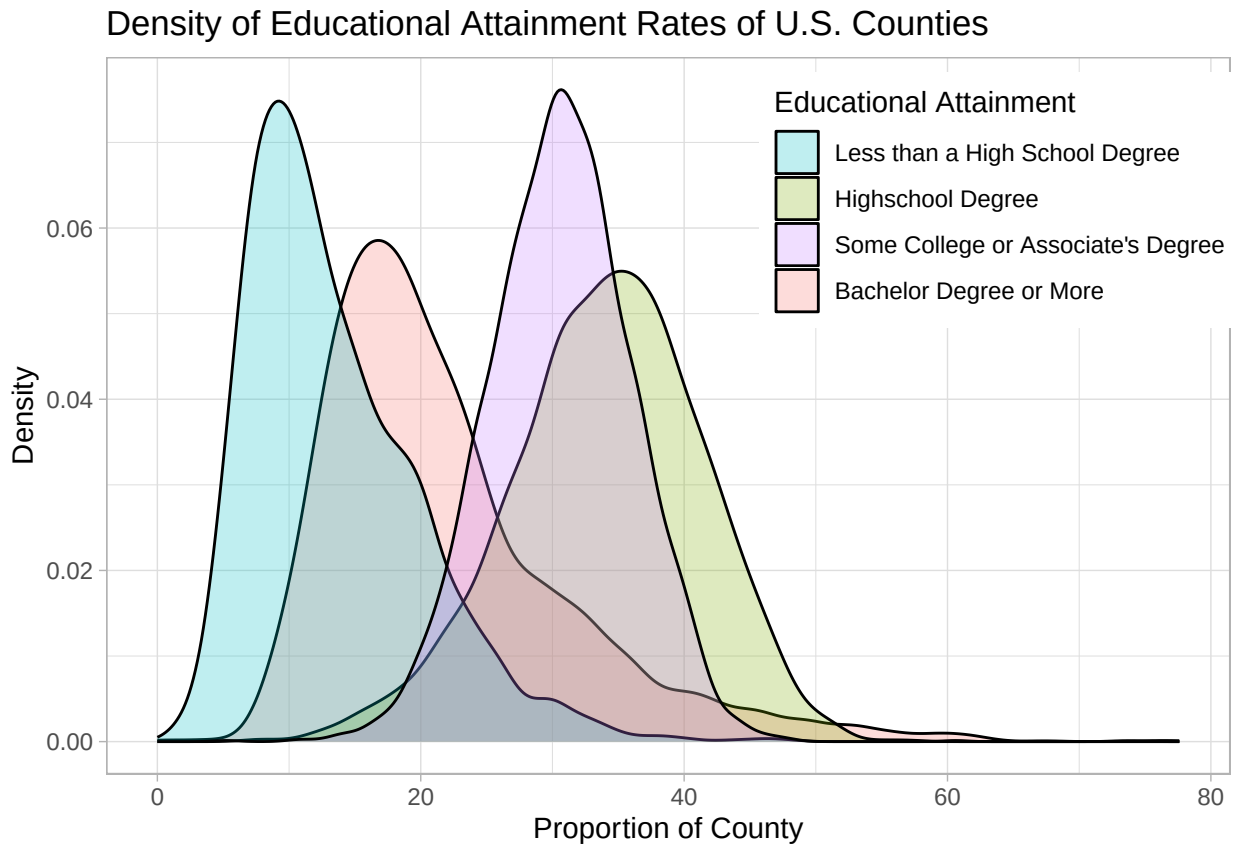
Poverty Rates



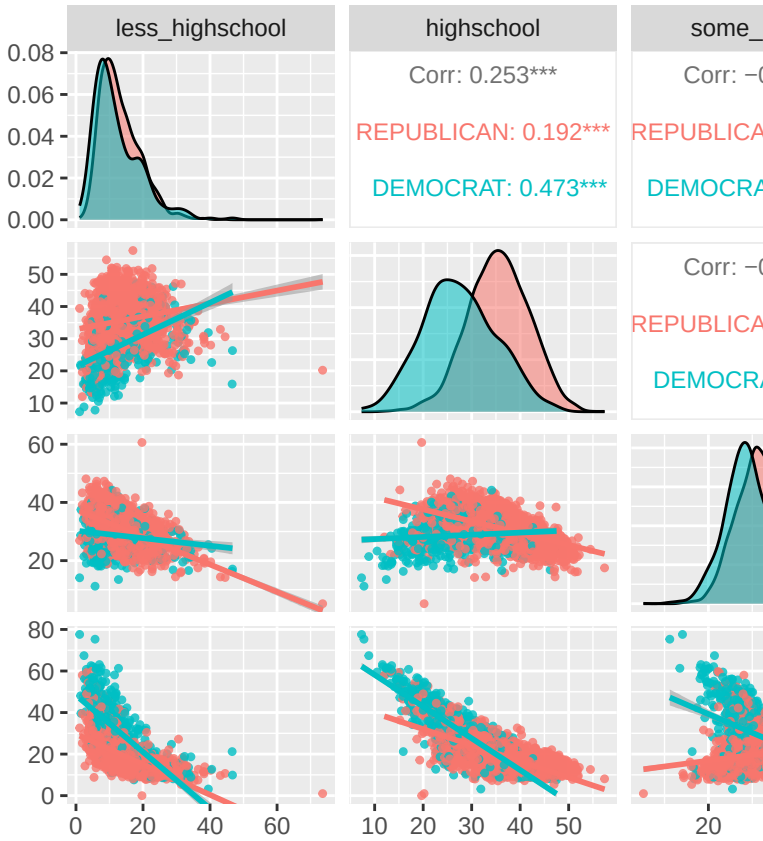
Density Curves of Poverty Rates in 2019

Above, the density curve of the poverty rate shows a peak around 12% for about a little over 8% of U.S. Counties. Most U.S. counties appear to have a poverty rate less than 25%, and there appear to be some outlying counties with over a 40% poverty rate. Separating the poverty rate by the winning party of the county shows that the poverty rate in democratic voting counties has a wider spread than the republican voting counties despite the republican voting counties peaking slightly higher than their democratic counterparts. For republican voting counties the peak unemployment rate appears to be around 12% for 10% of republican voting counties, while the peak unemployment rate for democratic voting counties is around 10% for about 6% of democratic voting counties. Counties with a greater than 40% poverty rate seem to almost be entirely comprised of democratic voting counties.

Educational Attainment Rates



Looking at the spreads of different levels of educational attainment overlapped shows how they are distributed. For those with less than a high school diploma the distribution appears skewed right and peaks at a rate of around 9% for about 7.5% of counties. Those with a high school diploma seems to have a centered distribution around 35% of the county population having a high school diploma for about 5.5% of counties. The spread of those with a bachelor degree or higher appears to have the widest spread of the different educational levels as the distribution appears to range from about 5% to 70% of the county population. The peak for this distribution is around 16% of the county population having a bachelor degree or higher for about 6% of counties.



Educational Attainment Rates by Winning Party

The density curves appear to very similar between republicans and democrats for those who do not have a high school degree and those who have some college or and associate's degree. On average it appears that counties who voted republican have a higher proportion of the population with a high school degree. In counties that voted democrat, there appears to be a greater spread in the proportion of the population that has a bachelor's degree or higher, while counties that voted republican are centered around 15% of the population having a bachelor's degree or higher.

When comparing between different educational attainment levels, most appear to be negatively correlated, which makes sense as the higher the proportion for one education level in a county, the lower the others have to be. There appear to be a few exceptions, like the slight positive correlation between some college and bachelor's or more for counties that voted republican. There is also a positive correlation between the proportion of a county with less than a high school degree and with a high school degree regardless of the party the county voted for. A possible explanation is counties with a low proportion of people who did not graduate high school, but have a high proportion of people who graduate high school are driving the correlation to be positive.