**Overview**

The objective is to find out what factors are correlated with the voting for the 2020 Presidential Candidates: Joe Biden and Donald Trump.

**Data**

The data provided is survey information. It contains 8280 observations and 59 variables. It also contains a data dictionary which is shown in table 1. The data is primarily categorical data with the exception of Age. There are missing variables in 51 of the 59 variables.

|  |  |
| --- | --- |
| **caseid** | **Case identifier** |
| weight\_pre | Full sample pre-election weight |
| weight\_post | Full sample post-election weight |
| sampvar | Full sample variance unit |
| varstrat | Full sample variance stratum |
| interest\_politics | How often does R pay attention to politics and elections? |
| interest\_campaigns | How interested in following campaigns? |
| state\_reg | Voter registration state |
| party\_reg | Party of registration |
| primary\_voter | Did R vote in a Presidential primary or caucus? |
| pol\_spectrum | Where does R place themselves on seven-point liberal-conservative scale? |
| party\_id | Party identification |
| party\_salience | How important is R's party identity? |
| gov\_trust | How often does R trust government in Washington to do what is right? |
| gov\_interests | Is government run by a few big interests or for the benefit of all? |
| gov\_waste | Does the government waste much tax monthly? |
| gov\_corrup | How much of the government is corrupt? |
| people\_trusted | How often can people be trusted |
| gov\_responsive | How much do elections make the government pay attention? |
| better\_economy | Which party would do a better job of handling the nation's economy? |
| better\_health | Which party would do a better job of handling health care? |
| better\_immigratino | Which party would do a better job of handling immigration? |
| better\_taxes | Which party would do a better job of handling taxes? |
| better\_environment | Which party would do a better of handling the environment? |
| better\_covid | Which party would do a better job of handling COVID-19? |
| fav\_deathpen | Do you favor or oppose the death penalty? |
| stayhome | The country would be better off if we just stayed home |
| trust\_media | How much do you trust the news media |
| corr\_trump | How have government corruption levels changed since Trump was elected? |
| fav\_impeach | How much did you support the House impeachment of the President? |
| fav\_senacquittal | How much did you support the Senate's decision to acquit President Trump? |
| covid\_gov | How would you characterize the federal government's response to COVID-19 |
| covid\_reopen | Is America opening too quickly or too slowly in response to COVID-19? |
| inc\_gap | How much larger is the income gap today? |
| gov\_climate | Show the government do more about rising temperatures? |
| samesex | Should businesses be required to provide services to same sex couples? |
| transgender | What are your feelings on transgender bathroom use? |
| lgbtlaw | Do you favor laws that protect LBGT against job discrimination |
| birthright | Do you favor ending birthright citizenship |
| deportkids | Should children brought illegally be sent back or allowed to stay? |
| wall | Do you support building a wall on the border with Mexico? |
| russianinterfere | How likely was it that Russia will interfere with the upcoming election? |
| religion | Major religious group |
| age | Respondent's age in years |
| marital | Marital Status |
| education | Educational Attainment |
| spouse\_edu | Spouse's educational attainment |
| armedforces | Did R participate in US armed forces? |
| labor | Labor force status |
| union | Is anyone in HH part of a labor union |
| ethnicity | R's self-identified ethnicity |
| children | How many children aged 0 - 17 in household |
| income | Total household income |
| health | Health of R |
| getcovid | Did anyone in the household get COVID? |
| satisfied | How satisfied is R with life? |
| vote | Did R vote for President in 2020? |
| whovoted | For whom did R vote? |
| region | Census region |

**Table 1: Data Dictionary**

**Approach**

The analysis approach would be

* Review the data dictionary to get more context and select key variables
* Apply different techniques (visualization, statistical modeling) to the data in order to achieve the objective
* Generate insights

The analysis was done using R and RStudio.

**Analysis**

Multiple attributes were looked using the data dictionary. After the review it was narrowed down to 9 variables. These are:

* Ethnicity
* Income
* Education
* Marital status
* Religion
* Political Spectrum
* Party registration
* Age
* Whovoted

Univariate plots will be used to get a better understanding of these variables.

Looking at the whovoted variable it looks like Biden got slightly more votes than Trump. It did look competitve

**Figure 1: Whovoted plot**

Age skews towards an older demographic. 40+ and above.

**Figure 2: Age plot**

**Figure 3: Religion plot**

**Figure 4: Income plot**

**Figure 5: Ethnicity plot**

**Figure 6: Education plot**

**Figure 7: Party registration plot**

**Figure 8: Political spectrum plot**

**Figure 9: Martial status plot**

The data points are mainly categorical with the exception of Age. Therefore, to see if there is correlation between the different variables and whovoted the chi-square test will be applied. The null hypothesis of the Chi-Square test is that no relationship exists on the categorical variables. It would also use 95% confidence. All p-values are less than 0.05 hence we can conclude that the variables are not independent of each other and that there is a statistical relationship between the categorical variables.

Using bivariate plots …

**Figure 10: Age plot**

**Figure 11: Religion plot**

**Figure 12: Income plot**

**Figure 13: Ethnicity plot**

**Figure 14: Education plot**

**Figure 15: Party registration plot**

**Figure 16: Political spectrum plot**

**Figure 17: Martial status plot**

**Figure 18: Martial status plot**

**Insights**