Investigation of the Approval Ratings of COVID-10 Policies

due April 20, 2022 by 11:59 PM

Group 2-1: Kylie Azizzadeh, Hannah Lawry, Annabel Stickel, Nina Moske, Mele Buice

April 20, 2022

Load Packages

```
library(tidyverse)
library(tidymodels)
```

Load Data

```
covid_approval <- read.csv("covid_approval_polls_adjusted.csv")
```

Introduction and Data

Background and Significance:

According to the World Health Organization, the COVID-19 pandemic began in December of 2019 and has since killed over 6.2 million people globally. Approximately 80.6 million people have contracted COVID-19 in the United States since the pandemicbegan, and almost 988,000 of these people have died from the disease. In response to the United States government's response to COVID-19 - which involved lockdowns, mask mandates, and social distancing - three-quarters of the American public now saw that "the country is more divided than before the coronavirus outbreak," according to a Pew Research Center study. In the middle of the pandemic, anotherevent occurred that greatly impacted the United States: the 2020 Presidential Election between Joe Biden and incumbent Donald Trump. The pandemic and the election were not mutually exclusive. In fact, given the divisive nature of the pandemic and the already-divided political climate of the U.S., the pandemic played an important role in the 2020 election. Moreover, President Trump and President Biden carried out radically different plans for addressing COVID-19, prompting our interestin examining the impact of their approaches on their popularity.

Data:

```
## Rows: 3,191
## Columns: 19
## $ subject            <chr> "Biden", "Biden", "Biden", "Biden", "Biden", "Bide~
## $ modeldate          <chr> "2/17/2022", "2/17/2022", "2/17/2022", "2/17/2022"~
## $ party              <chr> "D", "D", "D", "D", "D", "D", "D", "D", "D", "D", ~
## $ startdate          <chr> "1/24/2021", "1/28/2021", "1/29/2021", "1/31/2021"~
## $ enddate            <chr> "1/26/2021", "2/1/2021", "2/1/2021", "2/2/2021", "~
## $ pollster           <chr> "YouGov", "Quinnipiac University", "Morning Consul~
## $ grade              <chr> "B+", "A-", "B", "B+", "B", "B+", "B+", "B+", "A-"~
## $ samplesize         <dbl> 477.00, 333.25, 808.00, 484.00, 564.00, 336.00, 56~
## $ population         <chr> "a", "a", "rv", "a", "a", "a", "a", "a", "a", "rv"~
```

```
## $ weight              <dbl> 0.6285238, 0.6317152, 0.8337467, 0.5493243, 0.8883~
## $ influence           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ multiversions       <chr> "", "", "", "", "", "", "", "", "", "", "", "", ""~
## $ tracking            <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ approve             <dbl> 84.00, 93.00, 89.00, 88.00, 89.22, 88.00, 89.00, 9~
## $ disapprove          <dbl> 3.00, 5.00, 7.00, 7.00, 7.14, 8.00, 5.00, 5.00, 2.~
## $ approve_adjusted    <dbl> 86.66765, 91.98234, 90.73956, 90.66765, 89.71046, ~
## $ disapprove_adjusted <dbl> 2.681704, 5.525958, 5.876871, 6.681704, 6.637609, ~
## $ timestamp           <chr> "09:13:08 17 Feb 2022", "09:13:08 17 Feb 2022", "0~
## $ url                 <chr> "https://docs.cdn.yougov.com/ld46rgtdlz/econTabRep~
```

The dataset was taken from a FiveThirtyEight dataset posted by Jasmine Mithani, Aaron Bycoffe, Christopher Groskopf, and Dhrumil Mehta. The data was collected between March of 2020 and February of 2022 using many different pollsters like YouGov, American Research Group, and Morning Consults. These polls were distributed to Americans across the United States and from different population groups,including adults, registered voters, and likely voters. Different pollsters used different sample populations and sample sizes.

The "Covid_Approval" dataset makes 3,191 observations of 19 variables. We plan to use the variables "subject", which states which Presidents' COVID-19 policiesthe poll was evaluating, "party", which states which political party the participantsare affiliated with, "approve", which tells us what percent of polled participants approved of the COVID-19 policies, "disapprove", which tells us what percent ofpolled participants disapproved of the COVID-19 policies, "samplesize", which tells us the sample size that each poll used, and "population", which states that population group surveyed in each poll. We also created a variable called "political party" which just alters the name of a polls' political affiliation to include the full word "Democrat", "Republican", or "Independent" for clarity.

Research Questions:

When conducting our analysis, our team sought to understand whether an individual's political affiliation to a party had an impact on their overall approval rating of Biden's COVID-19 policies. Moreover, we want to investigate if we can categorize an average Democrat's approval rating versus an average Republican's. We are also interested in assessing how independent variables like sample size, population, and party affects the overall approval rating of Biden's COVID-19 policies. Lastly, we aim to understand if Trump or Biden had an overall higher approved COVID-19 policy and to therefore deduce how much that approval played into the decision of the winning candidate, Biden.

Research Question 1: Do Democrats favor Biden's COVID-19 policies more than Republicans? Because the pandemic became such a deciding factor of policy during the election, we would predict that those affiliated with Biden's political party (Democrats) would favor his COVID-19 policies more than Republicans. Moreover, we would assume that they also voted for him.

Null Hypothesis: Democrats do not favor Biden's COVID-19 policies more than Republicans, and therefore, their means of approval ratings are equal.

Alternative Hypothesis: Democrats favor Biden's COVID-19 policies more than Republicans, and therefore, Democrats a have higher mean of approval ratings.

$H_0 : \mu_{DemocratsApproval} = \mu_{RepublicansApproval}$ $H_A : \mu_{DemocratsApproval} > \mu_{RepublicansApproval}$

Research Question 2: Does the pollster's overall grade, the sample size, and population taking the poll affect Biden and Trump's approval rates? These independent factors should have no impact over Biden and Trump's COVID-19 approval ratings as the polls represent individuals who, regardless of their voting status, have likely formed their own opinions on the pandemic.

Null Hypothesis: Pollster's overall grade, the sample size, and population taking the poll do not impact Biden's approval ratings.

Alternative Hypothesis: Pollster's overall grade, the sample size, and population taking the poll impact Biden's approval ratings.

$H_0 : beta_k = 0$ (each predictor has no effect on approval ratings) $H_A : beta_k ! = 0$ (each predictor has an effect on approval ratings)

Research Question 3: Did Biden or Trump have higher approval rates? We would predict that Biden had higher overall approval rates because he was elected president and won the election against Trump.

Null Hypothesis: Biden and Trump had equal overall approval ratings.

Alternative Hypothesis: Biden had higher overall approval ratings than Trump.

$H_0 : \mu_{BidenApproval} = \mu_{TrumpApproval}$ $H_A : \mu_{BidenApproval} > \mu_{TrumpApproval}$

Methodology

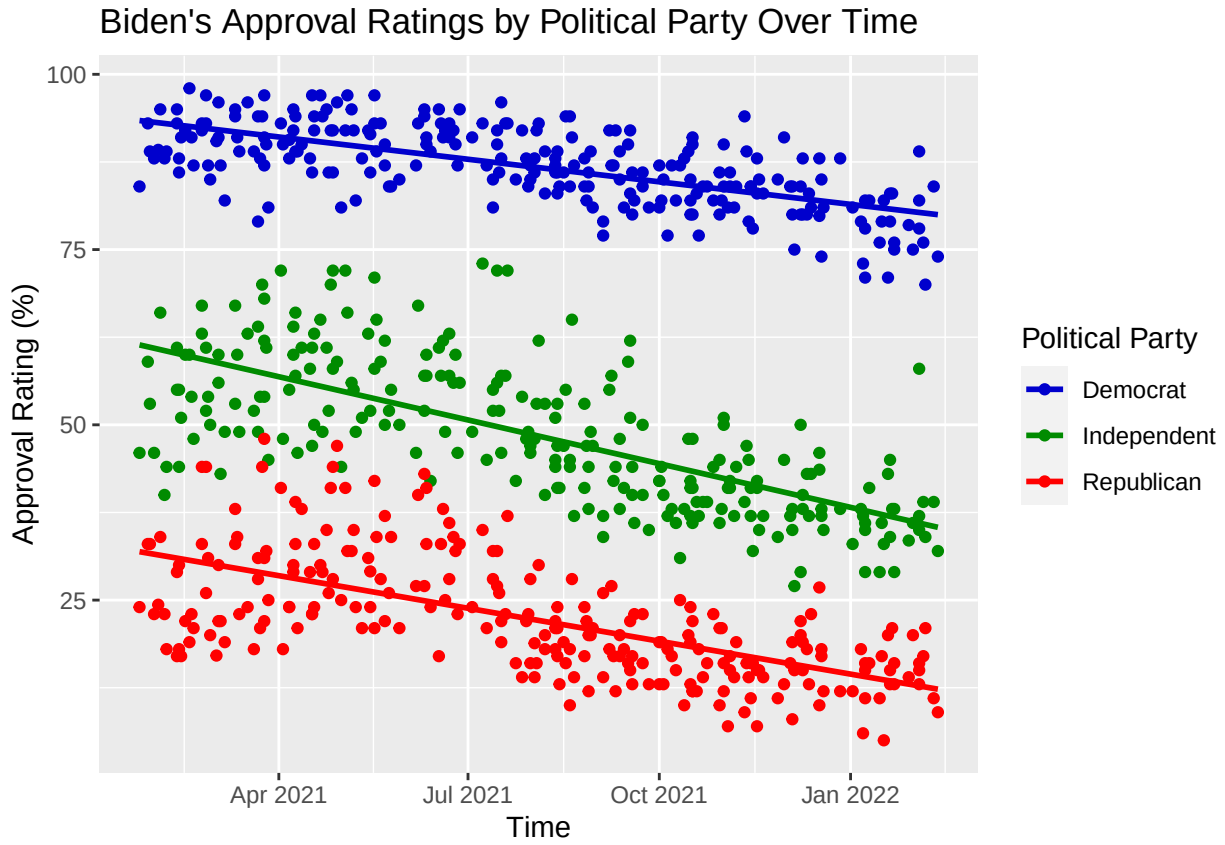
To begin our methodology, we will create a new variable called "political party" which specifically indicates whether the participants responding to a poll are Democrat, Republican, Independent, or of Mixed political parties.

Next, because our first hypothesis looks at only approval ratings of Biden's COVID-19 policies, we will create a new dataset with only the data from the polls addressing Biden's COVID-19 policies. Our second hypothesis looks at only approval ratings of Biden's COVID-19 policies and then only the approval ratings of Trump's COVID-19 policies, so we also made a dataset for Trump.

Research Question 1:

1. Scatterplot To start the investigation, we will look at a scatterplot to get an initial glance at the approval ratings of Biden's COVID-19 policies over time separated by political parties.

```
## `geom_smooth()` using formula 'y ~ x'
```



2. Probabilities We will define a "high approval rating" for Biden's COVID-19 policies as an approval rating over 52.97731 because 52.97731 is the average approval rating for Biden's COVID-19 policies. Therefore, we will also define a "low approval rating" for Biden's COVID-19 policies as an approval rating less than 52.97731.

To continue with our investigation of our first hypothesis, we will conduct some probability calculations. First, we will calculate the probability of a poll having a high approval rating of Biden's COVID-19 policies to see the likelihood of a poll with a high approval rating.

Next, we will calculate the probabilities of a high rating based on the affiliated political parties of the poll participants - the probability of a high rating given the participants were Democrats, the probability of a high rating given the participants were Republicans, and the probability of a high rating given the participants were Independents. Because our hypothesis predicts that Democrats will favor Biden's COVID-19 policies more than Republicans or Independents, we want to see if the probability that a poll gave Biden's policies a higher rating is different given the participants identified as Democrats, Republicans, or Independents.

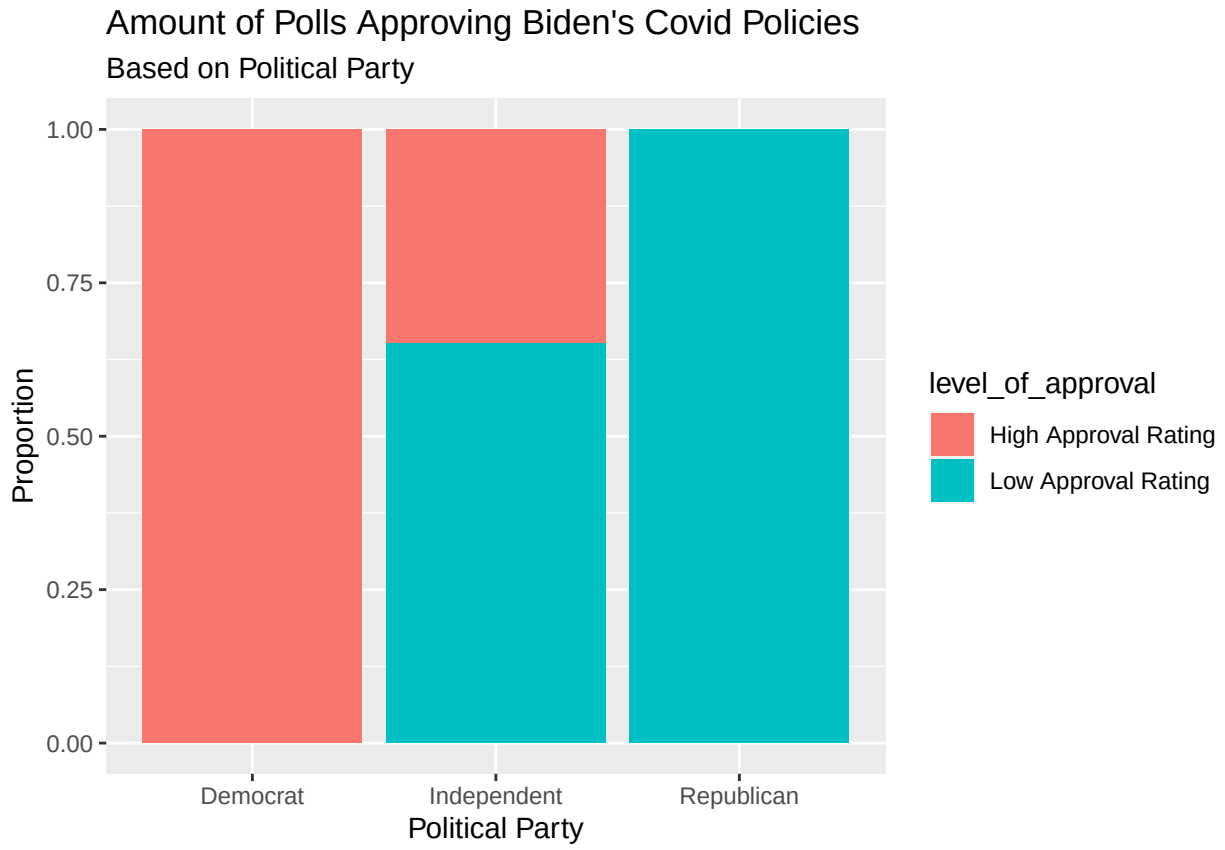
```
##   approve > 52.97731   n      prop
## 1              FALSE 558 0.5299145
## 2               TRUE 495 0.4700855
```

```
##   approve > 52.97731   n prop
## 1               TRUE 249    1
```

```
##   approve > 52.97731   n prop
## 1              FALSE 249    1
```

```
##   approve > 52.97731   n      prop
## 1              FALSE 159 0.6516393
## 2               TRUE  85 0.3483607
```

3. Bar Chart of Approval Ratings We will now create a bar graph to give us a visual of the proportion of polls with a high approval of Biden's COVID-19 policies versus a low approval of Biden's COVID-19 policies, separated by political party. To do this, we created a variable for the level of approval, with an approval rating above 52.97731 as a "high approval rating" and an approval rating below 52.97731 as a "low approval rating".



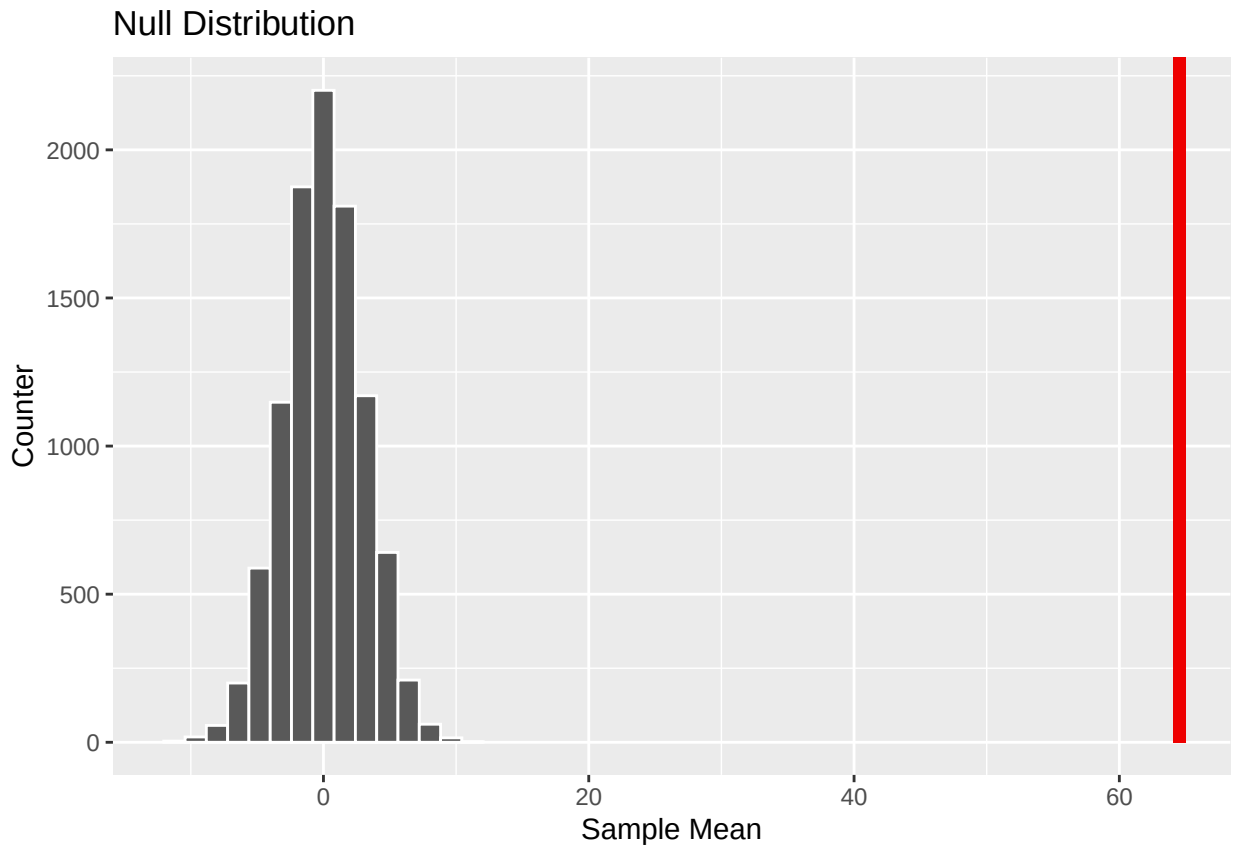
4. P-Value Next, we are calculating the p-value using a 0.05 significance level in order to asses whether we will reject or fail to reject the null hypothesis. Our observed statistic used is the difference in means of Democrat's and Republican's approval ratings for Biden's COVID-19 policy. Using this calculated p-value, we then visualize the null distribution.

```
##       mean na.rm
## 1 86.80759  TRUE
```

```
##       mean na.rm
## 1 22.26582  TRUE
```

```
## Warning: Please be cautious in reporting a p-value of 0. This result is an
## approximation based on the number of `reps` chosen in the `generate()` step. See
## `?get_p_value()` for more information.
```

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1       0
```



Research Question 2:

To investigate our second research question, we will create a linear regression with approval rating as the response variable and population, sample size, and party as the explanatory variables for Biden's COVID-19 policies, and then repeat the same for Trump's COVID-19 policies.

```
## # A tibble: 8 x 5
##   term           estimate std.error statistic   p.value
##   <chr>             <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)   52.7       0.748       70.5   0
## 2 populationlv   1.03      1.58         0.654 5.13e-  1
## 3 populationrv   2.51      0.523        4.80  1.80e-  6
## 4 populationv    7.02      8.29         0.848 3.97e-  1
## 5 samplesize     0.000161  0.000355     0.453 6.51e-  1
## 6 partyD        32.9       0.779       42.2   4.87e-228
## 7 partyI        -5.44      0.808       -6.73  2.77e- 11
## 8 partyR       -31.6       0.795      -39.8   1.29e-211
```

Of the explanatory variables investigated in terms of Biden's approval ratings, those with alpha level significance less than 0.05 are population, specifically population of registered voters, and all three parties.

```
## # A tibble: 7 x 5
##   term             estimate  std.error statistic  p.value
##   <chr>               <dbl>      <dbl>     <dbl>    <dbl>
## 1 (Intercept)   42.0        0.243        172.    0
## 2 populationlv   1.20       0.615          1.96  5.04e- 2
## 3 populationrv   0.710      0.266          2.67  7.66e- 3
```

```
## 4 samplesize     0.00000130 0.00000898     0.145 8.85e- 1
## 5 partyD       -29.5        0.332        -88.8   0
## 6 partyI        -5.49       0.339        -16.2   1.01e-55
## 7 partyR        39.9        0.332        120.    0
```

Of the explanatory variables investigated in terms of Trump's approval ratings, those with alpha level significance less than 0.05 are population, specifically population of registered voters, and all three parties.

Research Question 3:

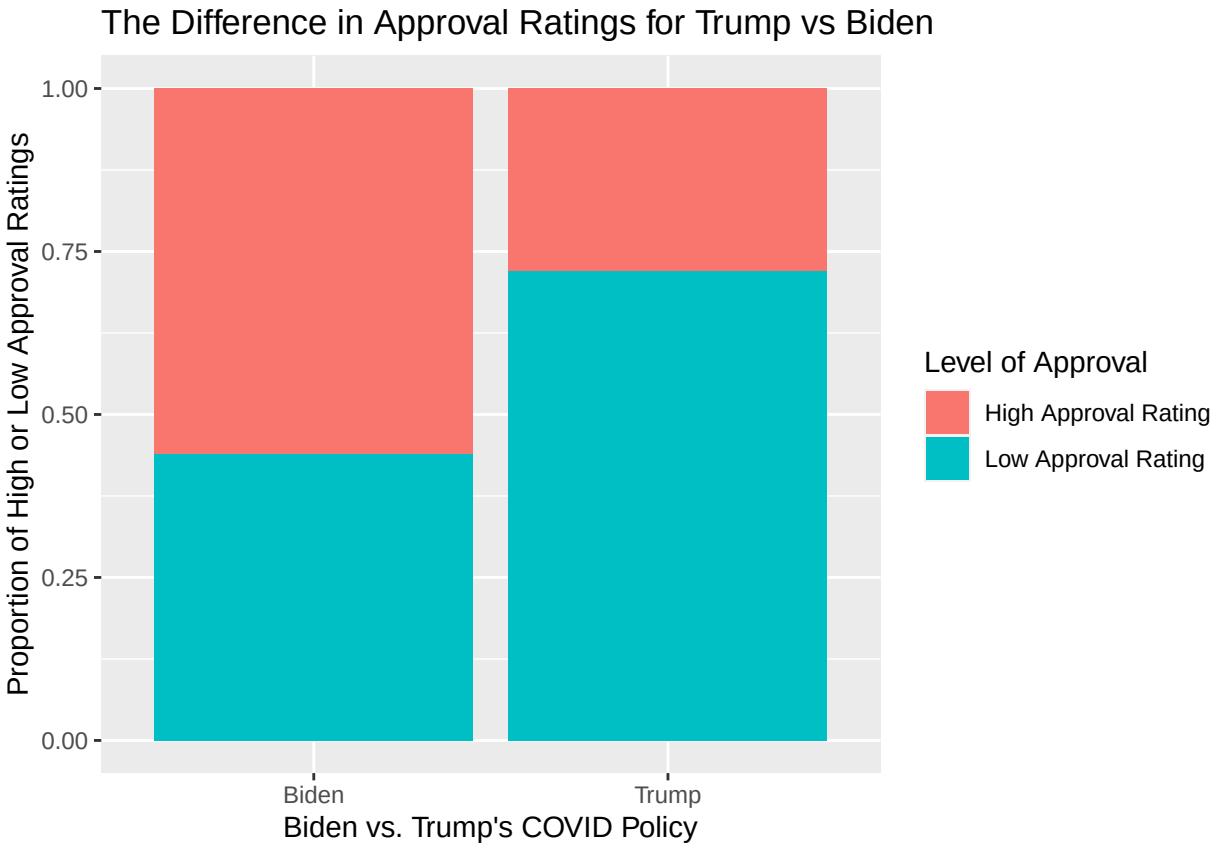
First, we are calculating the means of the approval rates of both Biden and Trump. We will then calculate the average of these means that will be used as a baseline for the high and low approval ratings.

```
##       mean na.rm
## 1 43.50299  TRUE
```

```
##       mean na.rm
## 1 52.97731  TRUE
```

```
##       mean na.rm
## 1 48.24015     1
```

Next, we are creating a bar graph to give us a visual representation of the proportion of high and low approval ratings, separated by Biden vs. Trump's COVID-19 policies. We added a variable for the level of approval, with an approval rating above 48.24015 as a "high approval rating" and an approval rating below 48.24015 as a "low approval rating". We got this number by averaging Trump and Biden's mean of approval ratings.



Results

Research Question 1: To start with our results, the scatterplot shows that across time, Democrats consistently had the highest approval rating, then Independents, and then Republicans. Over time, all of the lines of best fit initially slope up slightly, and then begin to slope downwards between April and July of 2021, and continue to slope down thereafter.

We then moved to calculating probabilities, rounding to the hundredth place. The probability that a poll gives Biden's COVID-19 policies a high approval (an approval rating above 52.97731) is 52.99%. The probability that a poll gave Biden's policies a high approval given the participants were Democrats is 100%. The probability that a poll gave Biden's policy a high approval given the participants were Republican is 0%. Additionally, the probability that a poll gave Biden's policy a high approval given the participants were Independent is 34.84%, though this statistic just adds context and doesn't necessarily add to our primary investigation.

Next, we used a bar graph to illustrate the proportions of polls that gave Biden's policies a high approval separated by party. The graph shows that the entire proportion of Democrat polls had a high approval, whereas the entire proportion of Republican polls had a low approval rating. Additionally, a little over 60% of Independent polls gave a low approval rating, while the rest gave a high approval rating. This visual makes sense given the probabilities we calculated, showing once again that all Democrats highly supported Biden's COVID-19 policies, whereas all Republicans did not.

In order to prepare for our null distribution where our statistic is the difference in means, we calculated the mean Biden approval ratings. For Democrats the mean approval rating was 68.61 and for Republicans the mean approval rating was 22.27.

Assuming that the alternative hypothesis is correct, the results of this visualization will show that the party an individual is affiliated with does not have significant weight on the approval rating this individual gives Biden's COVID-19 policy. We predict to see a bar chart graph that peaks in the middle, with the highest bar at the indifference level (b range) and potentially subtly right skewed with more values on the disapproval side than the approval side.

As for the results in finding and visualizing the p-value for research question 1, we calculated a p-value of 0, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis that there is no difference between democrats' and republicans' approval ratings of Biden's COVID policies. In the visualization, we can see that our observed statistic (red line) is significantly to the right of the sample means in the null distribution, which again shows that we should reject our null hypothesis.

Research Question 2: When analyzing Biden's approval ratings, the factors that affected his policy ratings regarding COVID-19 were the population of registered voters, and the affiliated party of each individual. Sample size had no apparent effect on his overall COVID-19 policy ratings, nor did the groups of likely voters, or adults. Rounding to the nearest hundredth decimal place, for each pollster polling a group of registered voters, Biden's COVID-19 approval ratings increased by 2.51%. For each pollster polling Democrats, Biden's COVID-19 approval ratings increased by 32.90%. For each pollster polling Republicans, Biden's COVID-19 approval ratings decreased by 31.63%. And for each pollster polling Independents, Biden's COVID-19 approval ratings decreased by 5.44%. The most weight was placed on Democrats and Republicans standing, so it makes logical sense that Democrats favored his policies and Republicans generally dissented. Moreover, their overall explanatory estimation of influence over approval ratings almost evenly cancels each other out.

When analyzing Trump's approval ratings, the factors that affected his policy ratings regarding COVID-19 were the population of registered voters, and the affiliated party of each individual. Similarly to Biden, sample size had no apparent effect on his overall COVID-19 policy ratings, nor did the groups of likely voters, or adults, although the likely voter population was much closer to our significance level than it was for Biden, indicating that party slightly favored Trump's policies. Rounding to the nearest hundredth decimal place, for each pollster polling a group of registered voters, Trump's COVID-19 approval ratings increased by 0.71%. For each pollster polling Democrats, Trump's COVID-19 approval ratings decreased by 29.51%. For each pollster polling Republicans, Trump's Covid approval ratings decreased by 39.94%. And for each pollster polling Indepedents, Biden's COVID-19 approval ratings decreased by 5.49%. The most weight was placed

on Democrats and Republicans standing, so it makes logical sense that Democrats favored his policies and Republicans generally dissented. Moreover, their overall explanatory estimation of influence over approval ratings almost evenly cancels each other out. Independents, in both cases, decreased the overall candidates approval ratings on COVID-19 policies.

Research Question 3: The results of our visualization for research question three concluded the support of our alternative hypothesis that Biden's COVID-19 policies garnered more support than Trump's did. We found the mean approval rating of Trump's COVID-19 policies to be 43.50, and the mean approval rating of Biden's COVID-19 policies to be 52.98, already indicating that Trump has a lower average approval rate. The bar chart exhibits that a little over 50% of participants gave Biden's COVID-19 policies a below-average rating, whereas about 75% of participants gave Trump's COVID-19 policies an above-average rating. This visual concludes that Biden's COVID-19 plans were more supported than Trump's.

Discussion

Overall, regarding research question 1, we learned that those who affiliate with the Democratic party are more likely to support Biden's COVID policies than those affiliated with the Republican party, showing our alternate hypothesis to be true. Our statistical analysis all pointed to the same results. Our statistical analysis included a scatterplot of Biden's approval ratings separated by party, a calculation of the probability of high approval ratings for Biden based on party, a bar graph with a visualization of the proportion of polls with a high vs. low approval of Biden's Covid-19 policies, a calculation of the p-value for the null hypothesis, and a visualization of the null distribution. These strategies were an appropriate statistical analysis because the graphs appropriately visualized the data and results for the question asked. Furthermore, using hypothesis testing, calculating the p-value, and null hypothesis visualization appropriately confirmed and supported the findings found from the graphs.

We also learned for research question 2 that the population of registered voters and affiliated party of each individual affected policy approval ratings; however, sample size and the group of likely voters did not. The statistical argument supporting these conclusions was our linear regression. This was an appropriate statistical analysis to answer this research question because linear regression is used to characterize the relationships between and impacts of different variables.

In the summary of our findings in research question 3, we learned that Biden generally had a higher proportion of high approval ratings than Trump did. Our statistical argument supporting our conclusions was our bar chart showing Biden's higher proportion of high approval ratings than Trump's, which was an appropriate analysis because it clearly visualized the answer to our research question. These statistics show that our third, final hypothesis was also true.

In general, these results were expected and made sense. The political party one is affiliated with generally influences their beliefs about what policies they support, and therefore, it makes sense that democrats would be in more support of the Democratic candidate's policies surrounding COVID-19. Furthermore, with Biden beating Trump in the election, it also is expected that he would have higher approval ratings than Trump, since COVID-19 was one of the most pressing and prevalent issues at the time of the election that would've influenced people's voting decisions.

As for a critique of our own methods, in research question 3, we could've added an analysis of the difference in means of the approval ratings for Biden and Trump, as well as a calculation of the p-value for the null hypothesis and an analysis of the null distribution. This could've helped confirm the significance of our results as shown by our bar graph. This is a general suggestion for improving our analysis and future work because by adding these complementary statistics, visualizations, and calculations, we are able to have additional support to the graph made, which gives a more comprehensive answer to the research question with more statistics backing our results. However, we thought the one bar graph was enough proof of our findings that Biden had a higher approval rating than Trump. Another critique for our project in general is that we could've added a wider variety of types of graphs, rather than just scatter plots and bar graphs.

Regarding the reliability and validity of our data, we think our data was extremely reliable. The data includes

results from multiple types of different Pollsters, which could prevent bias as opposed to if all the results were collected from one Pollster. Additionally, the Pollsters were rated using FiveThirtyEight's Pollster Ratings, as previously discussed in our introduction, so the data acknowledges which has more accuracy and a better methodology in collecting the data.

In terms of potential differences in our project if we were to start over, we would be open to adding and changing visualization and graphs depending on what best suits the data. We definitely would add more visual and statistical support of our research question number 3. It also would've been interesting to add a research question to see if Republicans favor Trump's policies more than Democrats, and then compare these results with that of research question 1. Furthermore, we could've also added a research question asking if Republicans favored Trump or Biden's COVID-19 policies more, as well as a question for that of Democrats.

References

- https://www.idea.int/sites/default/files/impact-of-covid-19-on-the-2020-us-presidential-election.pdf

- https://fivethirtyeight.com/features/how-much-did-covid-19-affect-the-2020-election/
- https://fivethirtyeight.com/features/whats-a-campaign-without-a-rally-how-coronavirus-could-alter-the-2020-election/
- https://www.pewresearch.org/2021/03/05/a-year-of-u-s-public-opinion-on-the-coronavirus-pandemic/
- https://www.pewresearch.org/fact-tank/2020/10/28/public-opinion-about-coronavirus-is-more-politically-divided-in-u-s-than-in-other-advanced-economies/
- https://covid19.who.int/

- https://www.usatoday.com/story/news/nation/2020/11/06/2020-election-american-divided-polarized-and-unsure-how-cope/6179404002/