ECON2250 Final Written Report

Dylan Hoge, Virthiha Selvamuthukumaran, Cordelia Tranfield

Dec 10, 2025

Introduction and data

Over the past decade, political polarization in the United States has intensified, with increasing divisions between the Democratic and Republican parties. As the U.S.'s political identities grow more entrenched, it is increasingly important to understand the factors that influence individuals' political affiliations. This research aims to explore the correlation between various key demographic factors—gender, race, and income—and political affiliation in the U.S. We aim to answer the question: How do gender, race, and income correlate with political affiliation (Democrat vs. Republican) in the United States?

The goal is to gain an understanding of how these factors impact voters' identification as either Democrats or Republicans. There is undoubtedly a growing need to understand the underlying drivers of America's political landscape. By examining the relationship between demographic characteristics and political affiliation, we can gain valuable insights into the voting behaviors and party loyalties that shape elections. This knowledge is essential not only for predicting future voting patterns but also for informing political campaigns, policy decisions, and even helping to mitigate the extreme division we see amongst politically active Americans today.

For our analysis, we pull from the 2020 American National Election Study (ANES) Time Series data set, which provides a comprehensive look at the political attitudes and behaviors of U.S. citizens. The data set contains responses from 5772 individuals, with 12 key variables relevant to the research question. After cleaning the data, the primary outcome variable was `party_dem`, which indicates whether respondents lean more toward the Democratic or Republican party. This variable was coded as 1 if a respondent leans Democrat (based on party ID categories 1–3) and 0 if they lean Republican (based on party ID categories 5–7). Pure independents or missing values are coded as N/A.

Gender was a binary variable indicating the respondent's gender (1 = female, 0 = male), and race was a binary variable indicating the respondent's race (1 = white, 0 = non-white). Our income categories, denoted as `income3` and `inc_sum`, categorized respondents into three income groups: low, medium, and high. `educ5` was a categorical variable representing the respondent's educational attainment, with five categories ranging from less than high school to graduate

or professional degrees (higher values represent more education). Finally, `ideology7` was a 7-point scale that asked respondents to place themselves on a political spectrum (1 = extremely liberal, 7 = extremely conservative).

From the descriptive statistics, we can begin to infer potential correlations between demographic factors and political affiliation. For instance, we may hypothesize that women and non-white individuals may be more likely to lean Democratic, while men and white individuals may lean Republican. Similarly, income and education levels are often strongly correlated with political party affiliation, with higher-income and more highly educated individuals typically aligning with the Republican party, while lower-income individuals with less education often identify more with the Democratic party.

This preliminary analysis lays the groundwork for further exploration into how gender, race, and income interact with political affiliation in the U.S. As we move forward, we conduct more sophisticated statistical analyses to test these relationships and draw conclusions about the influence of these demographic factors on political identity.

Methodology

Looking at the first hypothesis, the relationship between gender and party identification was analyzed. The null hypothesis is that the proportion of Democrats is the same for women as it is for men, or $\pi_{\text{women}} = \pi_{\text{men}}$. The alternative hypothesis is that women are more likely than men to be Democrats, or $\pi_{\text{women}} > \pi_{\text{men}}$. Due to the outcome being a numeric 0 or 1 and because we compare two independent groups, a two-sample t-test was used for the difference in mean between men and women. Since the sample size is very large, the sampling distribution of the mean is approximately normal when utilizing the Central Limit Theorem, so the t-test is the correct test to use even with this binary outcome. The assumptions made here are that the observations are independent of one another and that the two groups are random samples from the population.

Moving to our second hypothesis, the focus was on race and party identification. First, race was summarized with an indicator variable for being white, where `white` equals 1 for white respondents and 0 for non-white respondents. The null hypothesis for this states that the proportion of Democrats is the same for white and non-white respondents, or $\pi_{\text{white}} = \pi_{\text{nonwhite}}$. The alternative hypothesis for this case states that white respondents are less likely than non-white respondents to be Democrats, or $\pi_{\text{white}} < \pi_{\text{nonwhite}}$. Like the first hypothesis, a two-sample t-test was used when comparing the mean of `party_dem` for white and non-white respondents. The assumptions are also similar to the first hypothesis, assuming the observations are independent and that the sampling distribution of the mean is approximately normal. For descriptive purposes, we also made race into a three-category variable to distinguish white, Black, and Other to be represented by stacked bar charts of party identification by these groups.

The third hypothesis looks at income and party identification. Income is measured using the ANES family income categories. To classify this, a three-category income variable `income3` was created to classify respondents as low, middle, and high income, as well as a numeric scale, `inc_sum`, that keeps the ordered income categories. This hypothesis can be stated in two different ways. First, looking at just the group differences, the null hypothesis is that the proportion of Democrats is the same in the low, middle, and high income groups, or $\pi_{\text{Low}} = \pi_{\text{Middle}} = \pi_{\text{High}}$, and the alternative hypothesis is that at least one income group has a different proportion of Democrats. To test this hypothesis, a one-way analysis of variance, or ANOVA, was used, with `party_dem` as the outcome and `income3` as the factor. When using ANOVA, the assumptions made are the independence of observations, approximately normal residuals within each income group, and close to equal variances across groups. The second way, in terms of a trend, treats income as an ordered numeric predictor. The null hypothesis for this formulation is that there is no linear relationship between income and the probability of being a Democrat, so the slope for `inc_sum` is zero. The alternative hypothesis is that higher income is associated with lower probability of being a Democrat, showing a negative slope. This is tested through a simple linear regression of `party_dem` on `inc_sum`, which assumes linearity, independent errors, and approximately constant variance of residuals.

On top of these bivariate tests, a multiple regression model was used to look at all predictors together and provide context for each of the hypotheses. A linear probability model, or ordinary least squares regression, with `party_dem` as the outcome and gender, race, age, income, education, and ideology as the predictors, is specified as

$$Y = \beta_0 + \beta_1 \text{Female} + \beta_2 \text{White} + \beta_3 \text{Income} + \beta_4 \text{Age} + \beta_5 \text{Education} + \beta_6 \text{Ideology} + \varepsilon.$$

This model allows us to see whether the relationships between party identification and gender, race, and income persist after controlling for age, education, and political ideology.

Results and Discussion

In the cleaned ANES 2020 sample of 5772 respondents, a slight majority of the respondents identify as Democrat or lean Democrat, while the remainder identify as Republicans or lean Republican. Women constitute a small majority of the sample, whites are the largest racial group, and respondents are reasonably distributed across the three income categories, giving sufficient variation to evaluate our hypotheses.

Table 1: Gender Proportions

```
# A tibble: 2 x 4
  group     n prop_dem prop_dem_pct
  <chr> <int>    <dbl>        <dbl>
1 Men    2726    0.467         46.7
2 Women  3046    0.576         57.6
```
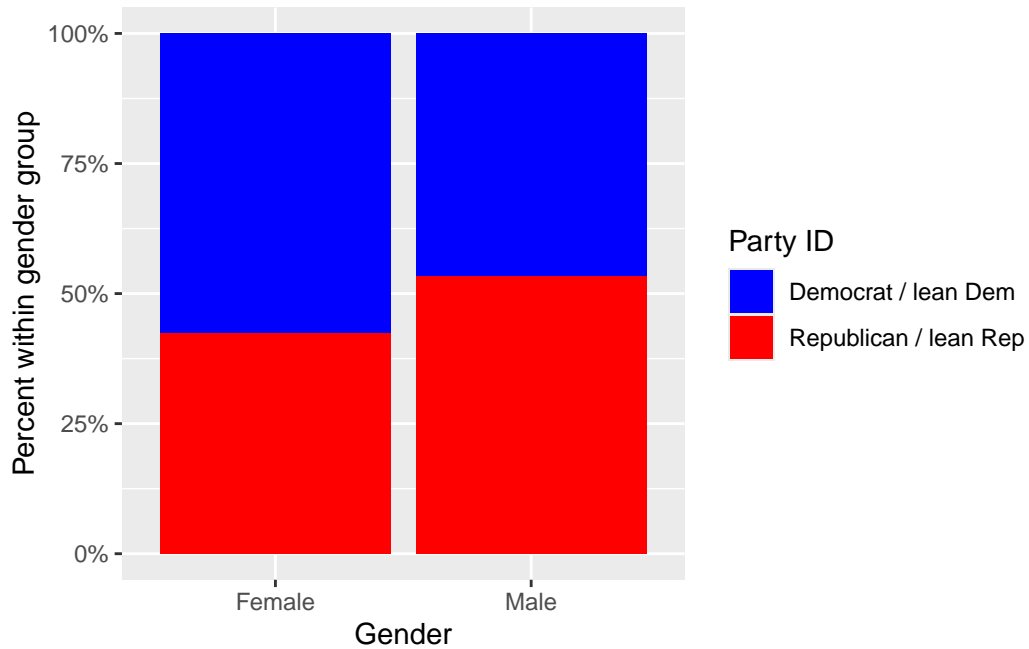

Figure 1: Party Identification by Gender

Gender had a significant link with party preference. Women were more likely to identify as Democrats than men, and this finding was supported by a very small p-value (7.7×10^{-17}). Although the percentage gap was not large, it was consistent enough to be statistically strong. This aligns with what is often seen in national elections: women tend to favor Democratic-leaning policies.

Table 2: Race Proportions

```
# A tibble: 2 x 4
  group         n prop_dem prop_dem_pct
  <chr>     <int>    <dbl>        <dbl>
1 Non-white  1354    0.716         71.6
2 White      4418    0.466         46.6
```
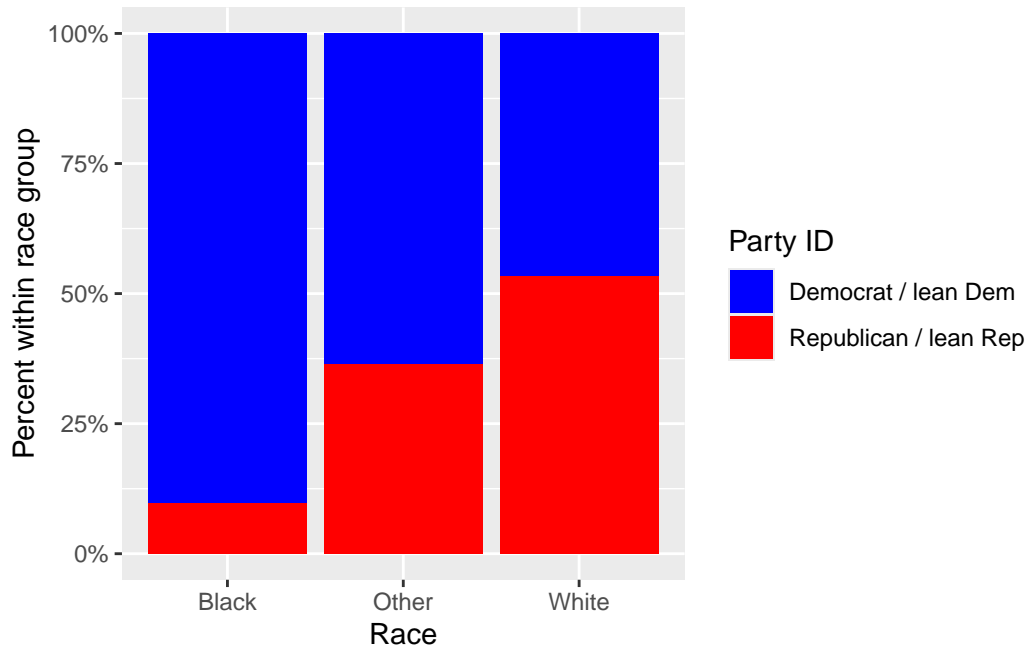

Figure 2: Party Identification by Race

Race produced the strongest and clearest relationship. Non-white respondents showed much higher rates of Democratic identification than white respondents, supported by an extremely small p-value (3.5×10^{-84}). This notable difference emphasizes how closely racial identity is tied to party identity in the U.S. Overall, race was the strongest predictor of political affiliation in our study.

Table 3: Income Proportions

```
# A tibble: 3 x 4
  income3      n prop_dem prop_dem_pct
  <fct>    <int>    <dbl>        <dbl>
1 Low       1493    0.557         55.7
2 Middle    1938    0.505         50.5
3 High      2341    0.521         52.1
```

Table 4: ANOVA

```
# A tibble: 1 x 6
  term        df sumsq meansq statistic p.value
  <chr>    <dbl> <dbl>  <dbl>     <dbl>   <dbl>
1 income3      2  2.40   1.20      4.82 0.00812
```
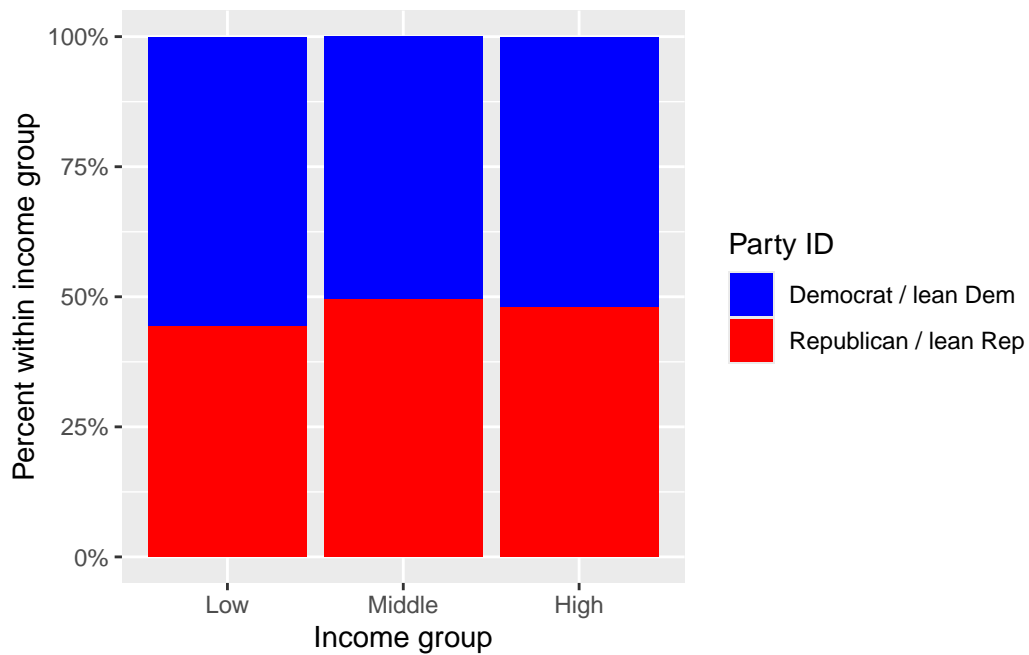

Figure 3: Party Identification by Income Group

Income was different. While income groups did show varying political affiliations, the differ-

ences were smaller. Lower-income individuals were more likely to lean Democratic compared to middle- and upper-income respondents, but the gap was not as large as the differences seen for gender or race. The ANOVA p-value of 0.00812 indicates that income does matter statistically, though not as strongly. This suggests that income affects political preferences, but it probably interacts with other factors like education, ideology, or region.

Table 5: Regression Model

```
# A tibble: 7 x 3
  term         estimate  p.value
  <chr>           <dbl>    <dbl>
1 (Intercept)    1.44   0
2 female         0.029 7.89e- 4
3 white         -0.159 4   e-53
4 inc_sum       -0.002 1.93e- 2
5 age            0.001 1.48e- 5
6 educ5          0.007 1.21e- 1
7 ideology7     -0.214 0
```

The multiple regression model helps to bring everything together to fully contextualize these findings by examining all predictors together. Controlling for race, income, age, education, and ideology, women remain about 3 percentage points more likely than men to identify as Democrats, shown in the coefficient of 0.029 for the `female` variable, which is still statistically significant. Also, white respondents are about twenty percentage points less likely than non-white respondents to be a Democrat, shown in the `white` coefficient of -0.195, one of the largest effects across the entire model. Higher income keeps be associated with lower probability of being a Democrat and the `inc_sum` is -0.002 with a statistically significant p-value barely, as the effect is small in magnitude. Age is also a factor, as the coefficient of -0.009 means that the older respondents are more Republican than younger respondents, even after controlling for the other variables. Education has a positive but weaker statistical association with Democratic identification, with a 0.007 coefficient that is not specifically different from zero at the significance levels. Over all the variables, ideology is by far the strongest predictor, as each step towards the conservative end of the seven point ideology scale decreases the probability of identifying as a Democrat by about 21 percentage points, as reflected by the -0.214.

When all is considered, the results show that gender, race, and income all matter for party identification in the United States, with ideology and race flexing their muscles with some very large notable effects. Women and non-white respondents are more likely to identify as Democrats, while higher income and older age are associated with a higher likelihood of identifying as a Republican. These quantifiable patterns provide strong support for the first two hypotheses and a more modest support for the third hypothesis, while the regression model

highlights that these relationships persist even when other important variables like education and ideology are taken into account.

Conclusion

Through this project, we gained valuable insight into how gender, race, and income relate to political affiliation in the United States. By looking at national survey data, we identified clear trends that explain why some groups lean more Democratic or Republican. Overall, the results strongly supported two of our hypotheses, while the third showed a smaller but still important relationship.

There were limitations to our analysis. Some of our variables were simplified, such as coding race as white vs. non-white, which does not capture the full diversity of racial categories in the U.S. Our analysis was also correlational, which means we can see relationships but cannot claim that demographics directly cause someone to vote a certain way. Despite these limits, our findings still provide useful insights. In the future, a more detailed statistical approach, such as logistic regression or multivariate analysis, could help isolate the independent effect of each factor.

Understanding these trends is increasingly crucial in a polarized political climate. These patterns influence how campaigns target voters, how policies appeal to different groups, and how Americans divide along party lines. While demographic traits are not the only forces driving political behavior, they remain strong indicators of party affiliation, and our analysis helps clarify why those divides persist today.