

# HonorsProject

Ariel Ooms

2025-04-05

1. A short explanation of your research question, and the reason why you selected certain data and variables.
2. A description of the main variables you are working with. A sentence or two for each variable should suffice.
3. Summary statistics
4. Descriptive analysis
5. Visualizations using ggplot2 (+ Any extra analysis of the data you find interesting) Each summary statistic result, descriptive analysis and plot or otherwise constructed result should include a short explanation of how that specific result relates to the research question you started with

## Research Question

My research question surrounds how different demographics and characteristics influence a person's party affiliation. I want to research if there is a correlation between one's identity and how they associate with politics. I selected my data because it has a lot of valuable information about differing partisan feelings, including on different scales. The variables I selected surround either political ideology or certain demographic factors so I can investigate if there is a relationship between the two.

## Variables

- educ - educ is the Education Level of the survey respondent. It includes people with no high school education, just a high school diploma, some college, 2-year college, 4-year college, and graduate school education levels.
- marstat - marstat is Marital Status of the survey respondent. Respondent answers included married, divorced, never married, separated, domestic/civil partnership, and widowed.
- faminc\_new - faminc\_new is the Income of the survey respondent's household or family. It included ranges from less than 10,000 to 500,000 or more, with several ranges in between.
- employ - employ is the Employment Status of the survey respondent. Answers included homemaker, retired, part-time, full-time, unemployed, student, permanently disabled, temporarily laid off, student, and other.
- child18 - child18 is the number of kids under 18 in the house of the survey respondent. This was a yes or no answer.
- pid3 - pid3 is the political ideology of the survey respondent based on the 3 party scale: democrat, republican, or independent
- pid7 - pid7 is the political ideology of the survey respondent based on the 7 party scale: strong democrat, lean democrat, not very strong democrat, independent, strong republican, lean republican, not very strong republican)

- ideo5 - ideo5 is a scale of the survey respondent's political ideology from very conservative, conservative, not sure, moderate, liberal, and very liberal.
- inputstate - inputstate is the state of residence of the survey respondent.
- democrat\_therm\_1 - democrat\_therm\_1 is the survey respondents feelings about democrats on a scale from 0-100
- republican\_therm\_1 - republican\_therm\_1 is the survey respondents feelings about republicans on a scale from 0-100

## Load in Data and Necessary Libraries

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)  
library(rnaturalearth)  
library(sf)
```

```
## Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

```
partisan <- read.csv("/Users/arielooms/Documents/College/Spring 25/PLS 202/partisan_data/s120-2025-week1.csv")  
#filter data down to just the variables I am interested in  
partisan <- partisan %>%  
  select(c(educ, marstat, faminc_new, employ, child18, pid3, pid7, ideo5, inputstate, democrat_therm_1, republica  
n_therm_1))
```

# Summary Statistics

```
head(partisan)
```

```
##          educ  marstat          faminc_new  employ child18          pid3
## 1   Post-grad  Married $100,000 - $119,999 Full-time    Yes  Republican
## 2     2-year  Married $100,000 - $119,999 Full-time    No   Not sure
## 3     2-year  Married  Less than $10,000 Homemaker    Yes  Republican
## 4     4-year  Divorced  $80,000 - $99,999 Full-time    No   Democrat
## 5     4-year  Married  $40,000 - $49,999  Retired     No  Independent
## 6 Some college Divorced  $40,000 - $49,999  Retired     No  Independent
##
##                pid7          ideo5 inputstate democrat_therm_1
## 1      Strong Republican Very conservative California          14
## 2                Independent          Not sure  Minnesota          50
## 3 Not very strong Republican Very conservative  Florida          13
## 4      Strong Democrat      Moderate  Indiana          99
## 5      Lean Republican      Moderate  Florida           0
## 6      Lean Democrat      Moderate  Virginia          25
##  republican_therm_1
## 1                94
## 2                50
## 3                99
## 4                 7
## 5                51
## 6                25
```

```
summary(partisan)
```

```
##      educ      marstat      faminc_new      employ
## Length:1000    Length:1000    Length:1000    Length:1000
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##
##      child18      pid3      pid7      ideo5
## Length:1000    Length:1000    Length:1000    Length:1000
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##
##      inputstate      democrat_therm_1      republican_therm_1
## Length:1000    Min.   : 0.00    Min.   : 0.00
## Class :character 1st Qu.: 21.00    1st Qu.: 18.00
## Mode  :character Median : 50.00    Median : 50.00
##                  Mean   : 50.25    Mean   : 47.53
##                  3rd Qu.: 79.25    3rd Qu.: 76.00
##                  Max.   :100.00    Max.   :100.00
```

Since most of the variables are not numeric, the only summary statistic that is relevant to my question is the distribution of the feelings on democrats and republicans. In this specific dataset, the average person leans slightly more toward democrat. The average feeling towards democrats is 50.25 compared to 47.53 on the republican scale. Additionally, feelings toward democrats is skewed toward the higher end of the scale.

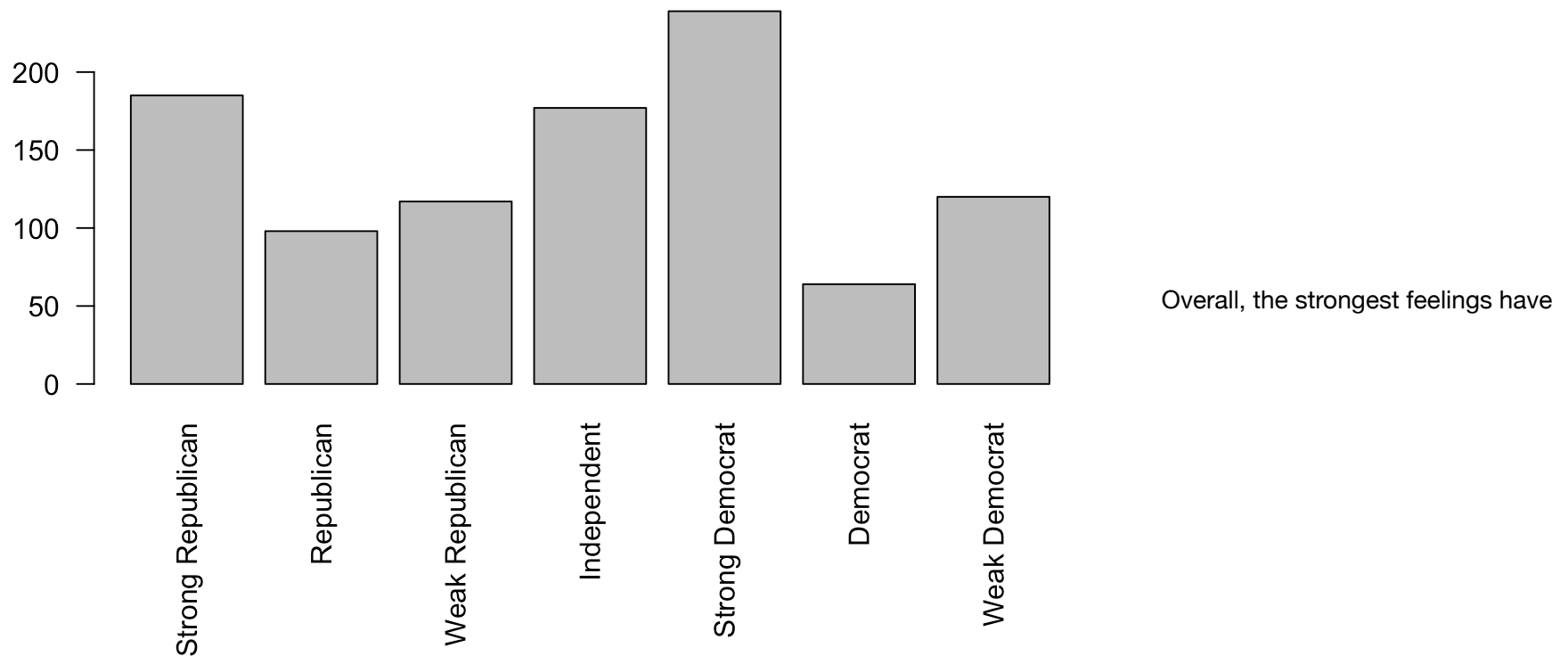
## Descriptive Analysis

I will start by simply visualizing the distribution of party affiliation across the United States as well as the distribution of participants in this survey.

```
#make party counts
strong_repub <- nrow(partisan[partisan$pid7=="Strong Republican",])
repub <- nrow(partisan[partisan$pid7=="Lean Republican",])
weak_repub <- nrow(partisan[partisan$pid7=="Not very strong Republican",])
independent <- nrow(partisan[partisan$pid7=="Independent",])
strong_dem <- nrow(partisan[partisan$pid7=="Strong Democrat",])
dem <- nrow(partisan[partisan$pid7=="Lean Democrat",])
weak_dem <- nrow(partisan[partisan$pid7=="Not very strong Democrat",])
party_counts <- c(strong_repub, repub, weak_repub, independent, strong_dem, dem, weak_dem)
names(party_counts) <- c("Strong Republican", "Republican", "Weak Republican", "Independent", "Strong Democrat",
"Democrat", "Weak Democrat")
party_counts
```

## Strong Republican	Republican	Weak Republican	Independent
## 185	98	117	177
## Strong Democrat	Democrat	Weak Democrat	
## 239	64	120	

```
# set margins - c(bottom, left, top, right)
par(mar = c(10, 4, 4, 2))
barplot(party_counts, las=2)
```



the highest frequencies. Strong democrat was the most frequent response, followed by strong republican. Independent was third most frequent. This indicates that people tend to have very strong opinions and fewer feel mildly about their political affiliations.

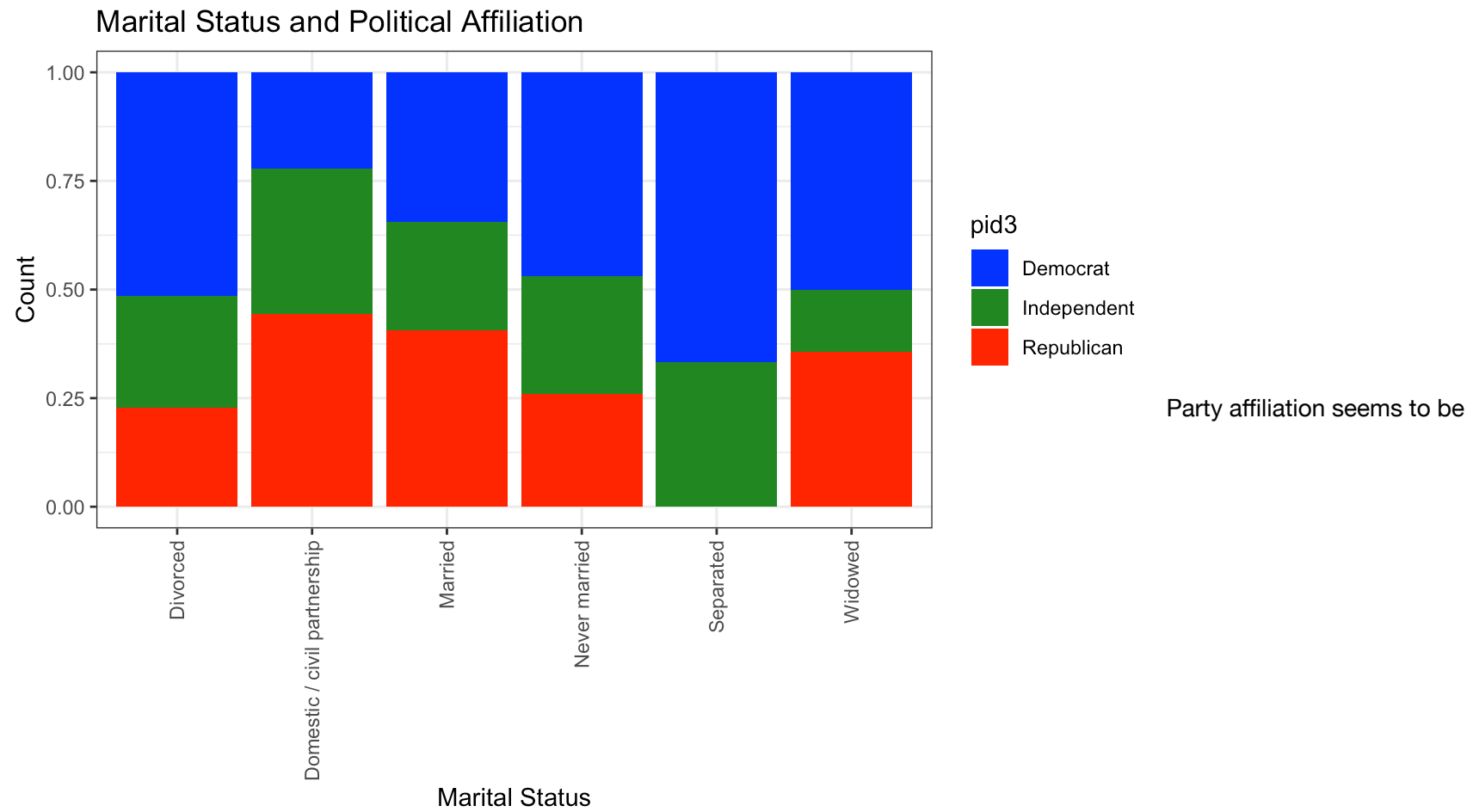
## Plot Visualization

Since most people lean strongly one way or another, I will focus on pid3 to make results easier to interpret. I want to look at how different factors affect the distribution of party affiliation. I am specifically looking at marital status, education level, and employment status. I want to exclude participants who answered "other" or "not sure" to better display the data.

```
partisan <- partisan %>%  
  filter(pid3==c("Democrat", "Republican", "Independent"))
```

```
## Warning: There was 1 warning in `filter()`.  
## i In argument: `pid3 == c("Democrat", "Republican", "Independent")`.  
## Caused by warning in `pid3 == c("Democrat", "Republican", "Independent")`:  
## ! longer object length is not a multiple of shorter object length
```

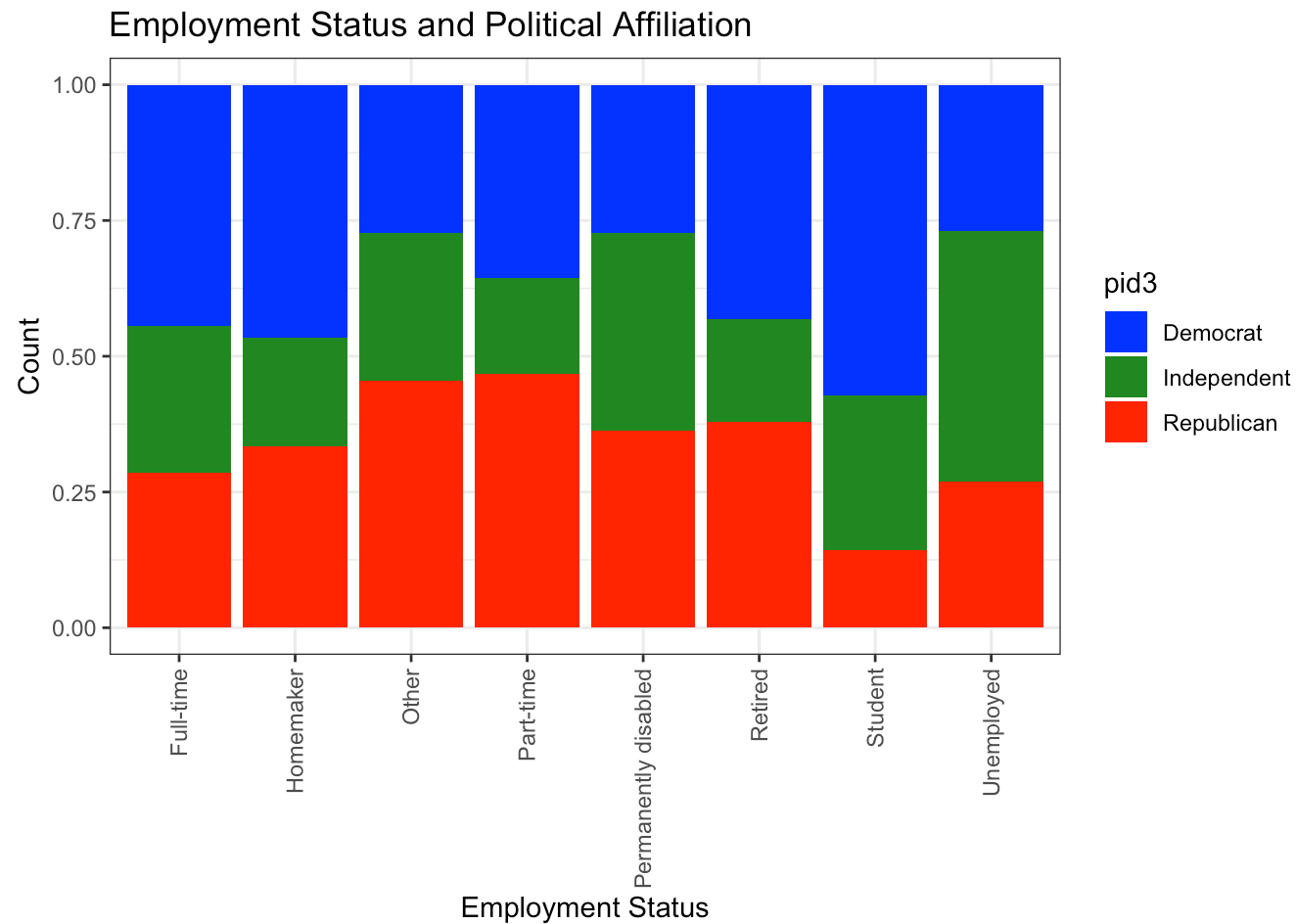
```
ggplot(data=partisan, aes(x=marstat, fill = pid3)) +  
  geom_bar(position = "fill") + labs(title="Marital Status") + theme_bw()+ scale_fill_manual(values = c("Republican" = "red", "Democrat" = "blue", "Independent" = "forestgreen")) +  
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) +labs(x="Marital Status", y="Count", title="Marital Status and Political Affiliation")
```



distributed the same no matter one's marital status. There does not appear to be an influence in one's political beliefs based on if they are married or not. There are some small variations however that stick out. For instance, no respondents that identified as separated reported themselves as a republican.

```
ggplot(data=partisan, aes(x=employ, fill = pid3)) +
  geom_bar(position = "fill") + labs(title="Employment Status") + theme_bw() + scale_fill_manual(values = c("Republican" = "red", "Democrat" = "blue", "Independent" = "forestgreen")) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) + labs(x="Employment Status", y="Count", title="Employment Status and Political Affiliation")
```

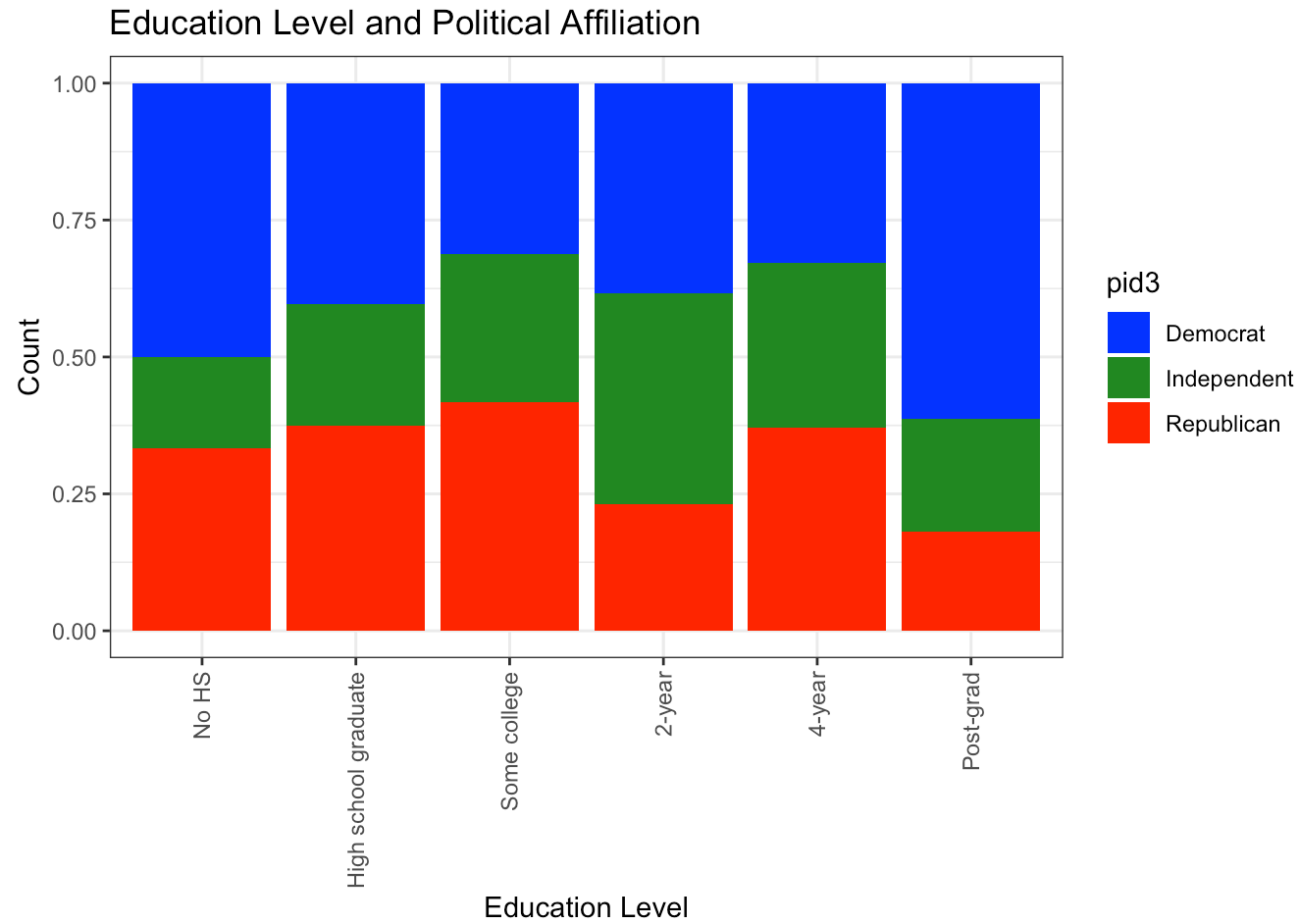




There are some interesting

associations between type of work and political affiliation. Most interesting to me was the large increase in democrat identifying participants when they are considered students. I also found the relationship between being unemployed and political party interesting as that is the group with the highest number of individuals reporting as independent. Beyond those two outliers, there does not seem to be an obvious relationship between work status and political opinion.

```
#change the order of the labels to be in ascending order of education
partisan$educ <- factor(partisan$educ, levels = c(
  "No HS",
  "High school graduate",
  "Some college",
  "2-year",
  "4-year",
  "Post-grad"
))
ggplot(data=partisan, aes(x=educ, fill = pid3)) +
  geom_bar(position = "fill") + labs(title="Education Status") + theme_bw() + scale_fill_manual(values = c("Republican" = "red", "Democrat" = "blue", "Independent" = "forestgreen")) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1)) + labs(x="Education Level", y="Count", title="Education Level and Political Affiliation")
```

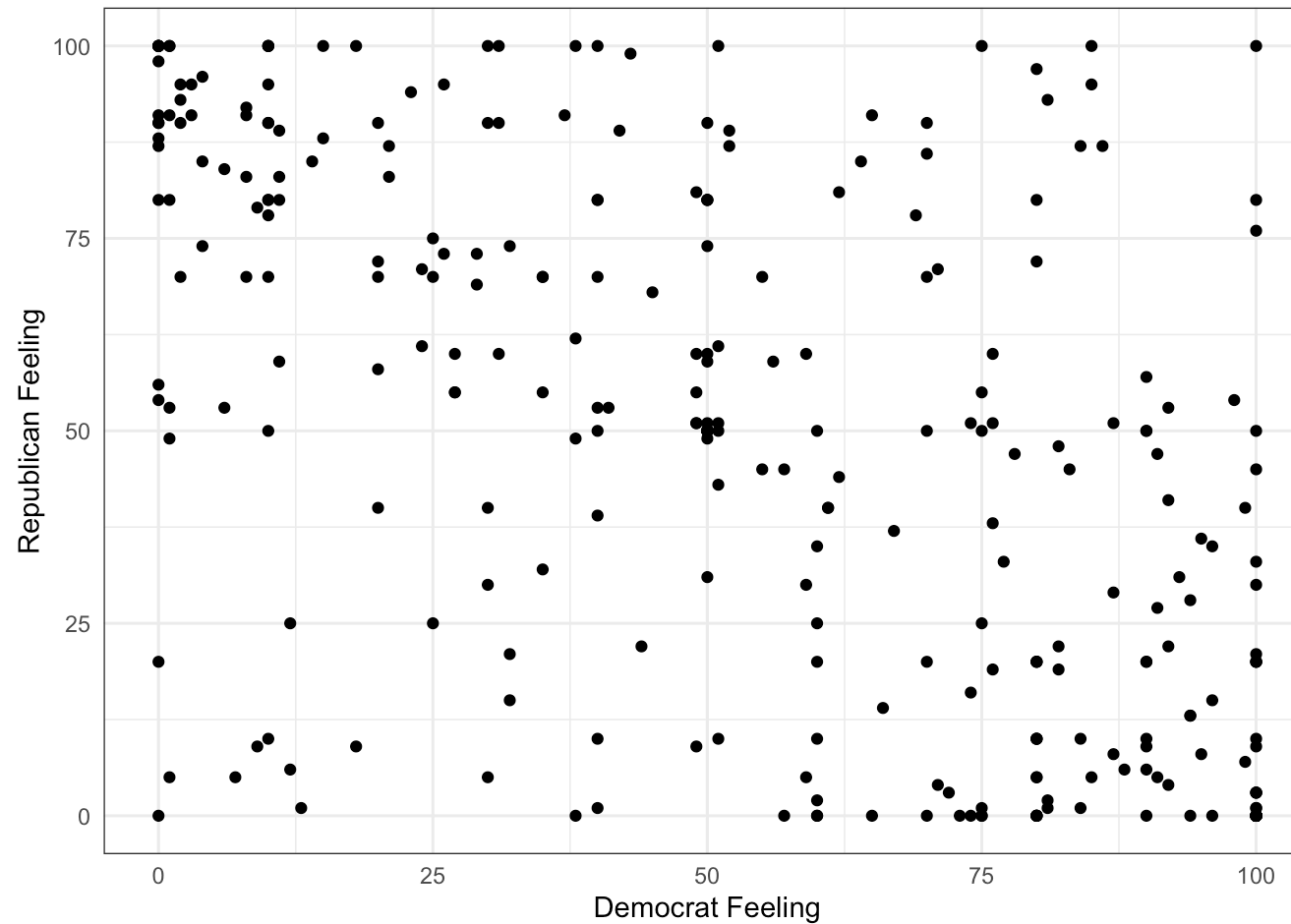


There does seem to be a trend with

political affiliation and education level. As education level increases, the amount of respondent's identifying as a democrat also increases. There is a very large jump between 4-year college and post-grad school where democrat identifying individuals skyrocket.

Lastly, I want to look into the relationship between `democrat_therm_1` and `republican_therm_1`. I will begin by simply plotting the two against each other.

```
ggplot(data=partisan, aes(x=democrat_therm_1, y=republican_therm_1)) + geom_point() + theme_bw() + theme(axis.ticks=element_blank(), legend.title=element_blank()) + labs(x="Democrat Feeling", y="Republican Feeling")
```



Just visually looking, the points are

very dispersed but there still appears to be a negative correlation. I will test the correlation and try to model a line of best fit to predict the feelings about republicans based on one's feelings toward democrats.

```
cor.test(partisan$democrat_therm_1, partisan$republican_therm_1)
```

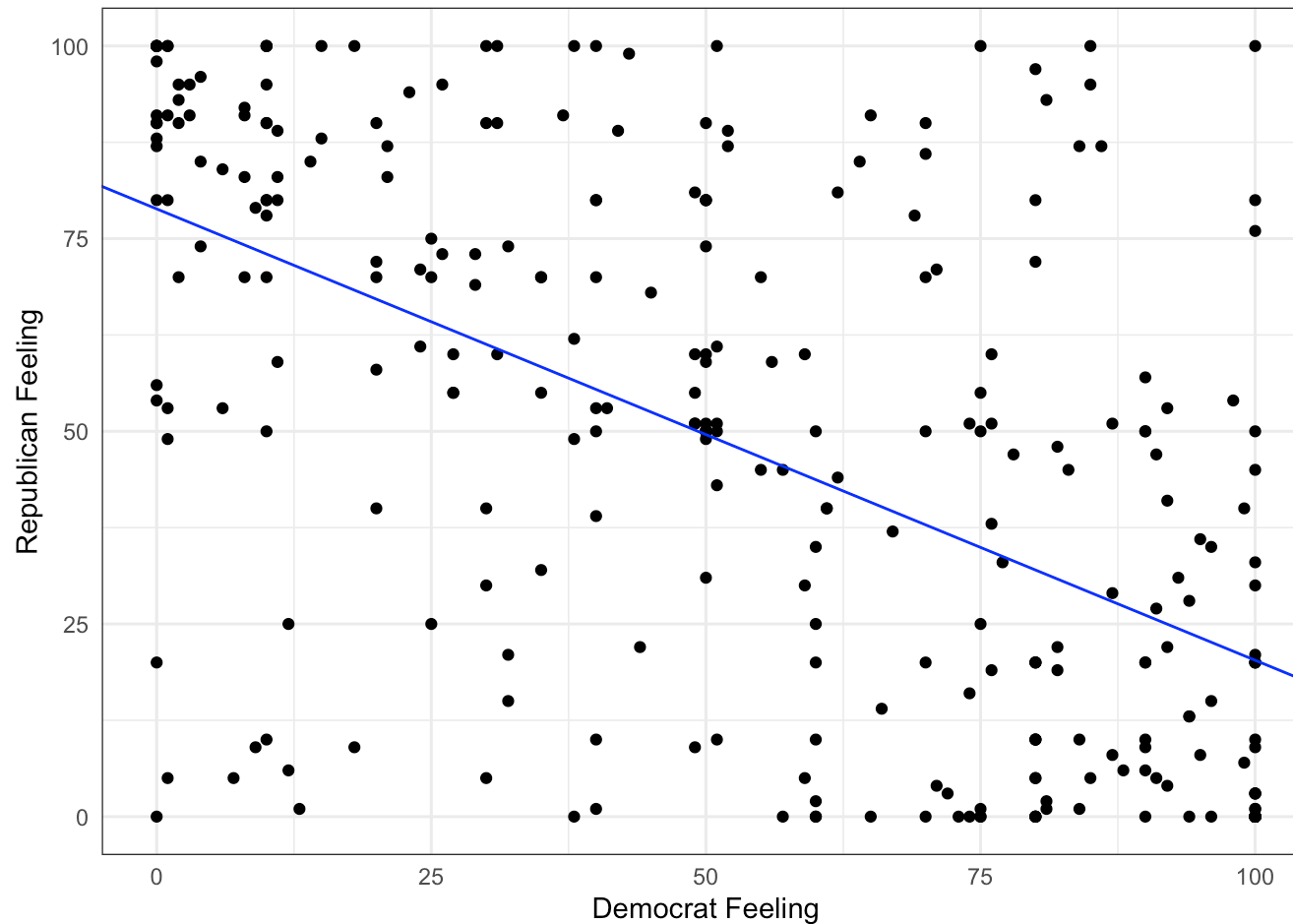
```
##
## Pearson's product-moment correlation
##
## data: partisan$democrat_therm_1 and partisan$republican_therm_1
## t = -11.957, df = 297, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.6419568 -0.4881692
## sample estimates:
## cor
## -0.5700349
```

The correlation test showed that the relationship between the two variables was very significant because the p-value (p-value < 2.2e-16) was extremely small, much smaller than 0.05, which is the threshold for significance.

```
dem_rep_model <- lm(partisan$republican_therm_1 ~ partisan$democrat_therm_1)
summary(dem_rep_model)
```

```
##
## Call:
## lm(formula = partisan$republican_therm_1 ~ partisan$democrat_therm_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -78.865 -20.289   0.423  17.564  79.711
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    78.86506     3.06571   25.73  <2e-16 ***
## partisan$democrat_therm_1 -0.58577     0.04899  -11.96  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.13 on 297 degrees of freedom
## Multiple R-squared:  0.3249, Adjusted R-squared:  0.3227
## F-statistic: 143 on 1 and 297 DF, p-value: < 2.2e-16
```

```
alpha <- coef(dem_rep_model)[1]
beta <- coef(dem_rep_model)[2]
ggplot(data=partisan, aes(x=democrat_therm_1, y=republican_therm_1)) + geom_point() + geom_abline(intercept=alpha, slope=beta, color="blue") + theme_bw() + theme(axis.ticks=element_blank(), legend.title=element_blank()) + labs(x="Democrat Feeling", y="Republican Feeling")
```



This model of best fit follows what the correlation test proved, that feelings toward democrats is negatively related to feelings toward republicans.

# Conclusion

Ultimately, the factors I looked into, marital status, education level, and employment status do not seem to share a strong relationship with political affiliation. Although there are a few interesting results and outliers, the actual spread of political ideology does not seem fully dependent on any one of these factors, and instead probably depends on multiple factors combined. However, one factor does seem to play a large role, which is feelings toward the opposite party of the respondent. If a person identifies high on the scale of feelings toward democrats, it is likely their score of feelings toward republicans will be low.