Gov 1372 - Hierarchies

Kayla Huang

September 30, 2021

Social Dominance and Politics

Welcome to the fourth Data Exploration Assignment. This week, we read about Social Dominance Theory, and its related psychological construct, Social Dominance Orientation (SDO). SDO is measured through a survey scale consisting of 16 items that we already explored in class. In this assignment, you will explore data on SDO and its relationship with other variables from a nationally representative survey fielded in 2018.

Note that the actionable part of each question is **bolded**.

Data Details:

- File Name: `sdo_data.csv`
- Source: These data are condensed and adapted from a survey by Data for Progress in 2018 (N = 3144). The data are representative of 2018 voters.

Variable Name	Variable Description
`sdo5`	Five-point social dominance orientation (SDO) scale: 1. Minimum SDO; . . . ; 5. Maximum SDO
`female`	Indicator for whether or not the respondent is female. Coded 1 if respondent is female, 0 otherwise.
`birthyr`	Respondent's birth year
`educ`	Education: 1. Didn't graduate HS; 2. HS graduate; 3. Some college; 4. 2-year college; 5. 4-year college; 6. Postgraduate degree
`race`	Race: 1. White; 2. Black or African-American; 3. Hispanic or Latino; 4. Asian or Asian-American; 5. Native American; 6. Mixed Race; 7. Other; 8. Middle Eastern
`favor_trump`	Favorability of Donald Trump: 1. very unfavorable; . . . ; 4. very favorable
`favor_blm`	Favorability of Black Lives Matter: 1. very unfavorable; . . . ; 4. very favorable
`favor_metoo`	Favorability of the Me Too movement: 1. very unfavorable; . . . ; 4. very favorable
`american_customs`	"The growing number of newcomers from other countries threatens traditional American customs and values": 1. Strongly disagree; . . . ; 5. Strongly agree
`race_ident`	"How important is being [respondent's race] to you?": 1. Not at all important; . . . ; 4. Very important

Variable Name	Variable Description
`pid3`	Three-category party identification: 1. Democrat; 2. Independent; 3. Republican
`ideo5`	Five-category political ideology: 1. Very liberal; . . . ; 5. Very conservative
`fear_of_demographic_change`	Fear of demographic change in the US: 0. Least fearful; . . . ; 1. Most fearful
`confederate_flag`	Is the Confederate flag mostly a symbol of slavery and white supremacy or Southern heritage and culture? Coded either "slavery" or "heritage"
`presvote16`	Vote choice in the 2016 presidential election

Question 1: REQUIRED

Before looking at data, the science of political psychology often involves building surveys. The teaching team builds the surveys you take using an online survey-building software called Qualtrics. This is often the same software that researchers use to build surveys and collect data that is eventually published in peer-reviewed journals. In this question, you'll create your own brief survey.

THIS QUESTION IS REQUIRED FOR ALL STUDENTS. Go to harvard.qualtrics.com and log in using your HarvardKey. Click "Create new project", then select "Survey". You can name your survey whatever you like. Leave the other two drop-down options at their default and click "Create project". Now you can input the SDO scale, which is given below. Make sure to include all 16 items, split into two sub-scales, in your survey. They are split into two sub-scales here, but they don't need to be in your survey. For each item, there should be seven response categories: Strongly favor, Somewhat favor, Slightly favor, Neutral, Slightly oppose, Somewhat oppose, Strongly oppose. Think about the format you think is best for these questions, available under "Question Type". How might the format of the questions affect the responses you get from the survey, or the experience respondents have while taking the survey? Also consider question ordering and how that may also affect the responses. BE SURE TO UPLOAD A SCREENSHOT OF YOUR QUALTRICS SURVEY TO YOUR BLOG THIS WEEK.

Dominance Sub-Scale

1. Some groups of people must be kept in their place.
2. It's probably a good thing that certain groups are at the top and other groups are at the bottom.
3. An ideal society requires some groups to be on top and others to be on the bottom.
4. Some groups of people are simply inferior to other groups.
5. Groups at the bottom are just as deserving as groups at the top.
6. No one group should dominate in society.
7. Groups at the bottom should not have to stay in their place.
8. Group dominance is a poor principle.

Anti-Egalitarianism Sub-Scale

1. We should not push for group equality.

2. We shouldn't try to guarantee that every group has the same quality of life.
3. It is unjust to try to make groups equal.
4. Group equality should not be our primary goal.
5. We should work to give all groups an equal chance to succeed.
6. We should do what we can to equalize conditions for different groups.
7. No matter how much effort it takes, we ought to strive to ensure that all groups have the same chance in life.
8. Group equality should be our ideal.

My survey can be found at the following link: https://harvard.az1.qualtrics.com/jfe/form/SV_egH0L3hQOFG4eNw.

In my survey, I created a matrix table. Here, survey respondents can see all the questions at once instead of considering every question in isolation. This may affect the results of my survey. I also considered the order in which questions appeared—for example, I selected the form to randomize the order in which the questions appear every time the survey is taken. In this way, we can try to avoid the situation in which a previous question has primed the individual to think in a certain way.

Question 2

Now let's take a look at the data.

```
sdo <- read_csv("sdo_data.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   sdo5 = col_double(),
##   female = col_double(),
##   age = col_double(),
##   educ = col_double(),
##   race = col_double(),
##   favor_trump = col_double(),
##   favor_blm = col_double(),
##   favor_metoo = col_double(),
##   american_customs = col_double(),
##   race_ident = col_double(),
##   pid3 = col_double(),
##   ideo5 = col_double(),
##   fear_of_demographic_change = col_double(),
##   confederate_flag = col_character(),
##   presvote16 = col_character()
## )
```

```
sdo
```

```
## # A tibble: 3,144 x 15
##     sdo5 female   age  educ  race favor_trump favor_blm favor_metoo
```

```
##    <dbl>  <dbl> <dbl> <dbl> <dbl>       <dbl>     <dbl>       <dbl>
##  1  3         1    75     2     1           4         1           1
##  2  2         0    73     5     1           4         1           2
##  3  1         0    69     2     1           1         4           4
##  4  2.5       0    79     6     1           4         1           2
##  5  1         0    64     5     1           1         4           4
##  6  1         0    60     3     7           1         4           4
##  7  1         1    49     6     2           1         4           4
##  8  3.5       0    66     5     1           4         1           1
##  9  3.5       0    76     6     1           4         1           1
## 10  1.75      0    38     6     1           1         4           4
## # ... with 3,134 more rows, and 7 more variables: american_customs <dbl>,
## #   race_ident <dbl>, pid3 <dbl>, ideo5 <dbl>,
## #   fear_of_demographic_change <dbl>, confederate_flag <chr>, presvote16 <chr>
```

What is the distribution of social dominance orientation in the sample? Make a plot, and report the mean and standard deviation of SDO in the sample. Extend this problem by splitting the plot by party ID of the respondent. Comment on what you find.

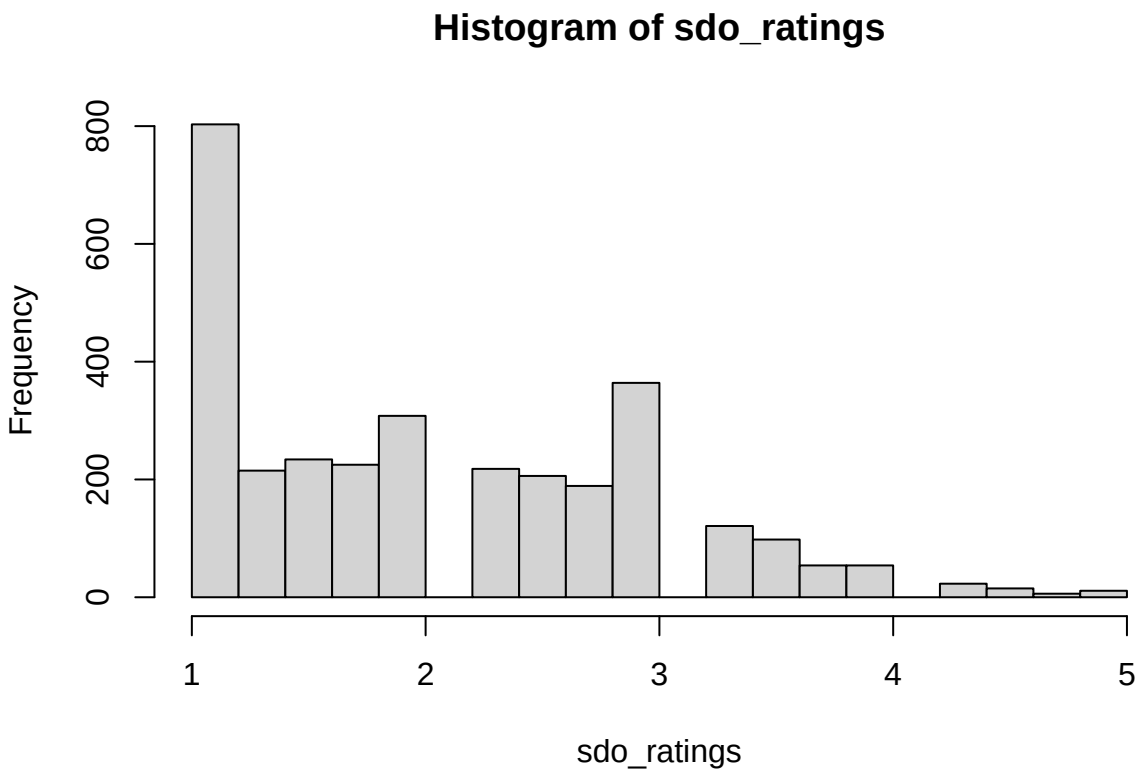
```
sdo_ratings <- sdo$sdo5
summary(sdo_ratings)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   2.000   2.053   2.750   5.000
```

```
sd(sdo_ratings)
```

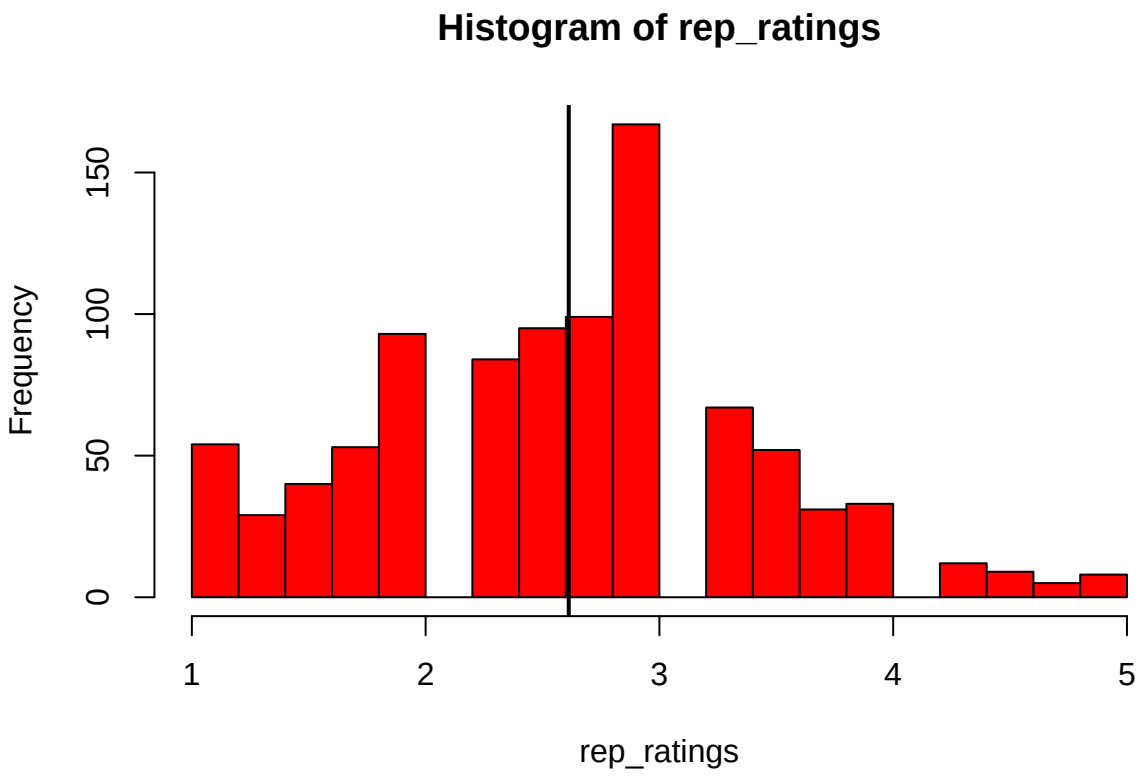
```
## [1] 0.9214715
```

```
hist(sdo_ratings, breaks=15)
```



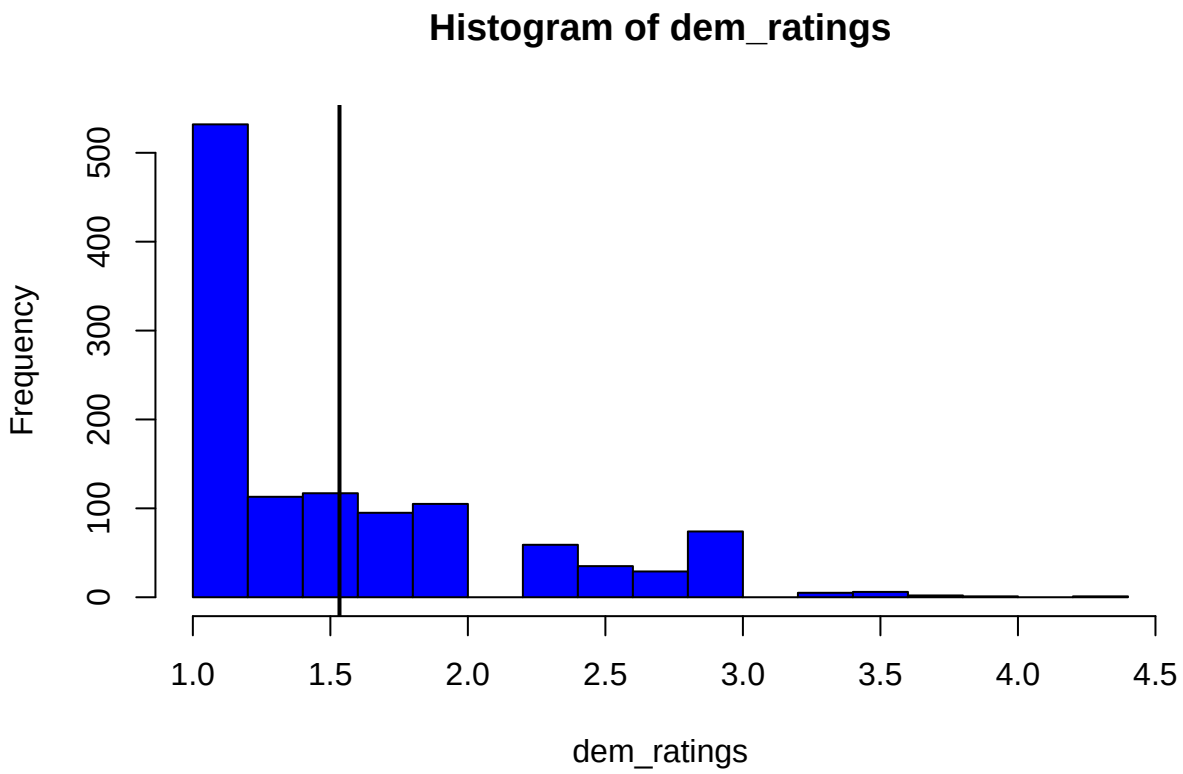
Now, splitting up this information by party, we see the following distribution for Republicans.

```
rep_ratings <- filter(sdo, pid3==3)$sdo5
hist(rep_ratings, breaks=15, col="red")
abline(v = mean(rep_ratings, na.rm = T), col = "black", lwd=2)
```



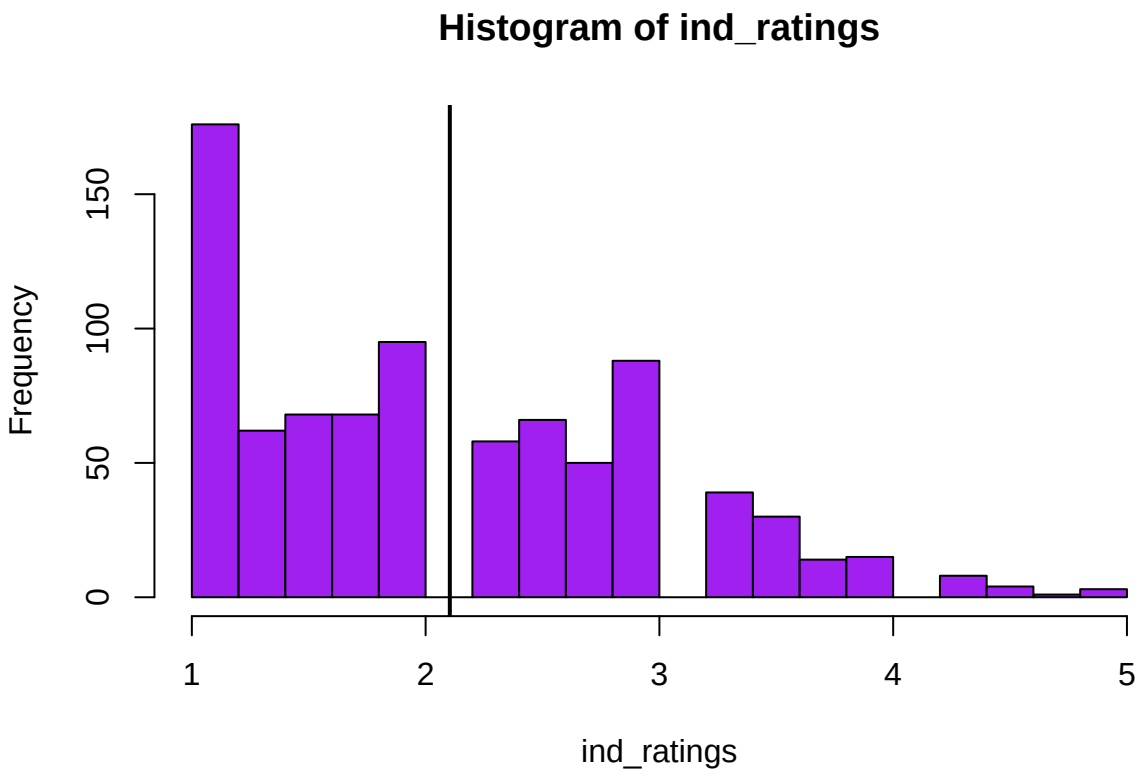
For Democrats:

```
dem_ratings <- filter(sdo, pid3==1)$sdo5
hist(dem_ratings, breaks=15, col="blue")
abline(v = mean(dem_ratings, na.rm = T), col = "black", lwd=2)
```



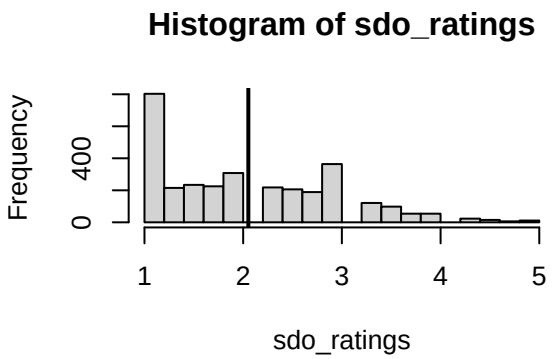
For independents:

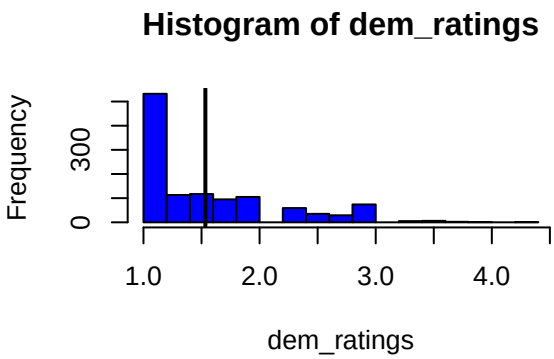
```
ind_ratings <- filter(sdo, pid3==2)$sdo5
hist(ind_ratings, breaks=15, col="purple")
abline(v = mean(ind_ratings, na.rm = T), col = "black", lwd=2)
```

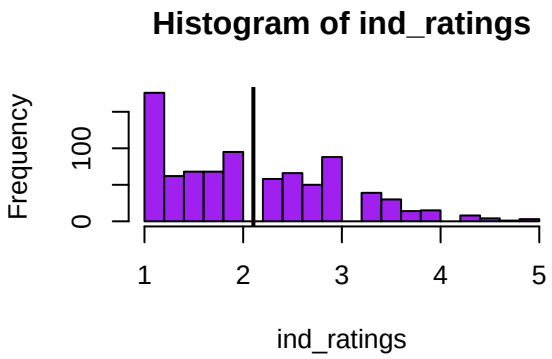


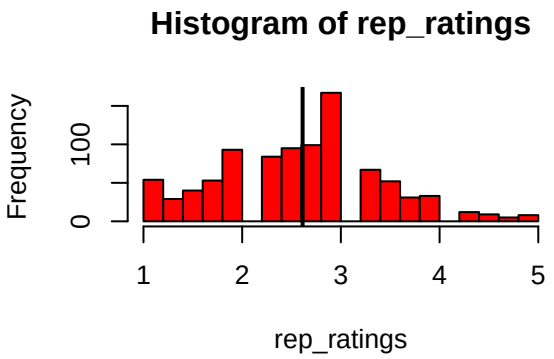
From these results, we can see, even just from the histogram, that Republicans tend to have a higher SDO than Democrats and Independents. Also, we see that Independents have a slighter higher SDO than Democrats as well, though not by much.

```
attach(sdo)
par(mfrow=c(2,2))
hist(sdo_ratings, breaks=15)
abline(v = mean(sdo_ratings, na.rm = T), col = "black", lwd=2)
dem_ratings <- filter(sdo, pid3==1)$sdo5
hist(dem_ratings, breaks=15, col="blue")
abline(v = mean(dem_ratings, na.rm = T), col = "black", lwd=2)
ind_ratings <- filter(sdo, pid3==2)$sdo5
hist(ind_ratings, breaks=15, col="purple")
abline(v = mean(ind_ratings, na.rm = T), col = "black", lwd=2)
rep_ratings <- filter(sdo, pid3==3)$sdo5
hist(rep_ratings, breaks=15, col="red")
abline(v = mean(rep_ratings, na.rm = T), col = "black", lwd=2)
```









Question 3

In the reading for this week, we saw that gender is central to social dominance theory, which predicts that men tend to have higher SDO than women do. **Is this true in this sample as well? Report the average SDO for men and women. Comment on what you find. Extend this question by reporting the difference in means along with the p-value. Is the difference significant at a .05 significance level?**

```
sdo %>% group_by(female) %>% summarize(mean(sdo5))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 2
##   female `mean(sdo5)`
##    <dbl>        <dbl>
## 1      0         2.19
## 2      1         1.94
```

Yes, I found this to be true in this sample, too. The men have a higher average SDO than the women in this survey. Now, to test the difference in means.

```
t.test(sdo$sdo5, sdo$female)
```

```
##
##  Welch Two Sample t-test
##
## data:  sdo$sdo5 and sdo$female
## t = 80.693, df = 4834.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.470773 1.544018
## sample estimates:
## mean of x mean of y
## 2.0525604 0.5451654
```

The p-value is less than 0.05, so the result is significant.

Question 4

What is the correlation between sdo5 and the favor_trump variable? Is the correlation statistically different from zero? You can use cor.test() for this question. Interpret what you find. If you want, extend this question by creating a scatterplot with the line of best fit to visualize the relationship. You can use geom_point() in the ggplot architecture for this.

```
cor.test(sdo$sdo5, sdo$favor_trump)
```

```
##
##  Pearson's product-moment correlation
##
## data:  sdo$sdo5 and sdo$favor_trump
## t = 41.246, df = 3111, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5713838 0.6168261
## sample estimates:
##       cor
## 0.5945795
```

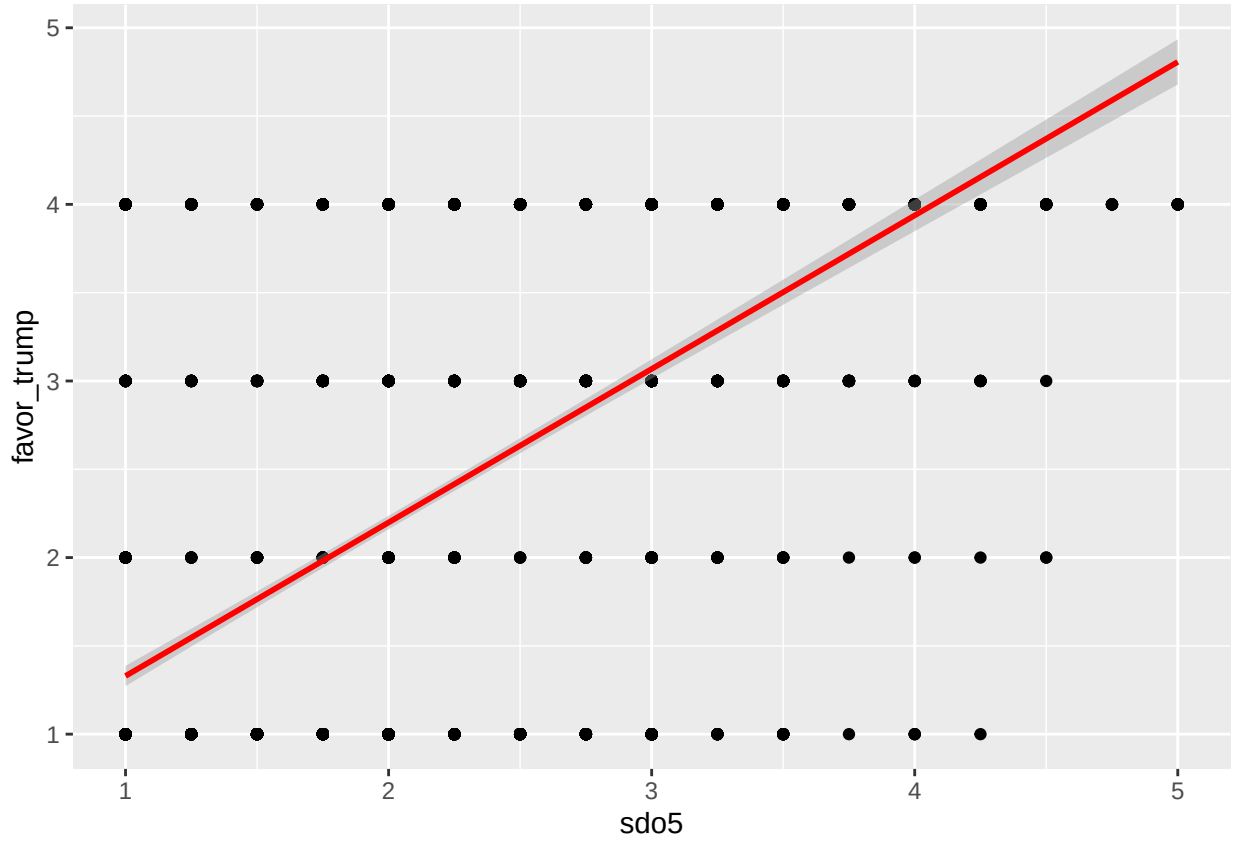
Since the p-value is definitely less than 0.05, these results are statistically significant.

```
ggplot(sdo, aes(sdo5, favor_trump)) + geom_point() + stat_smooth(method = "lm", col = "red")
```

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

```
## Warning: Removed 31 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 31 rows containing missing values (geom_point).
```



Question 5

Correlation matrices, like the one below, are useful for visualizing the pairwise relationships between several variables. They allow you to see the correlation coefficients of of many relationships at once. **Plot a correlation matrix of the correlation between SDO and some of the variables you think might be related to SDO and to each other. Choose at least 3 variables in addition to SDO. Before you make your plot, briefly discuss why you think the variables might all be related. The package `ggcorrplot` may be useful here. Discuss what you see in your plot.**

Here is an example of what your correlation plot might look like, using the `ggcorrplot` package:

```
# ggcorrplot
```

Question 6: Data Science Question

In this next question, we will use regression to model vote choice as a function of SDO and other variables of interest. This will help us get a fuller picture of the impact of social dominance orientation on political attitudes. We will fit the following model:

$$\text{rep\_vote} = \beta_0 + \beta_1 \text{sdo5} + \beta_2 \text{female} + \beta_3 \text{white} + \beta_4 \text{educ} + \beta_5 \text{age} + \beta_6 \text{pid3} + \beta_7 \text{ideo5} + \epsilon$$

You'll notice that the variable `white` doesn't exist in our data set. When doing regression analysis, researchers often code race as a binary - for example, 1 for white and 0 for all non-white. This is done largely to make the regression results easier to interpret. Without turning race into a binary variable, the regression model would

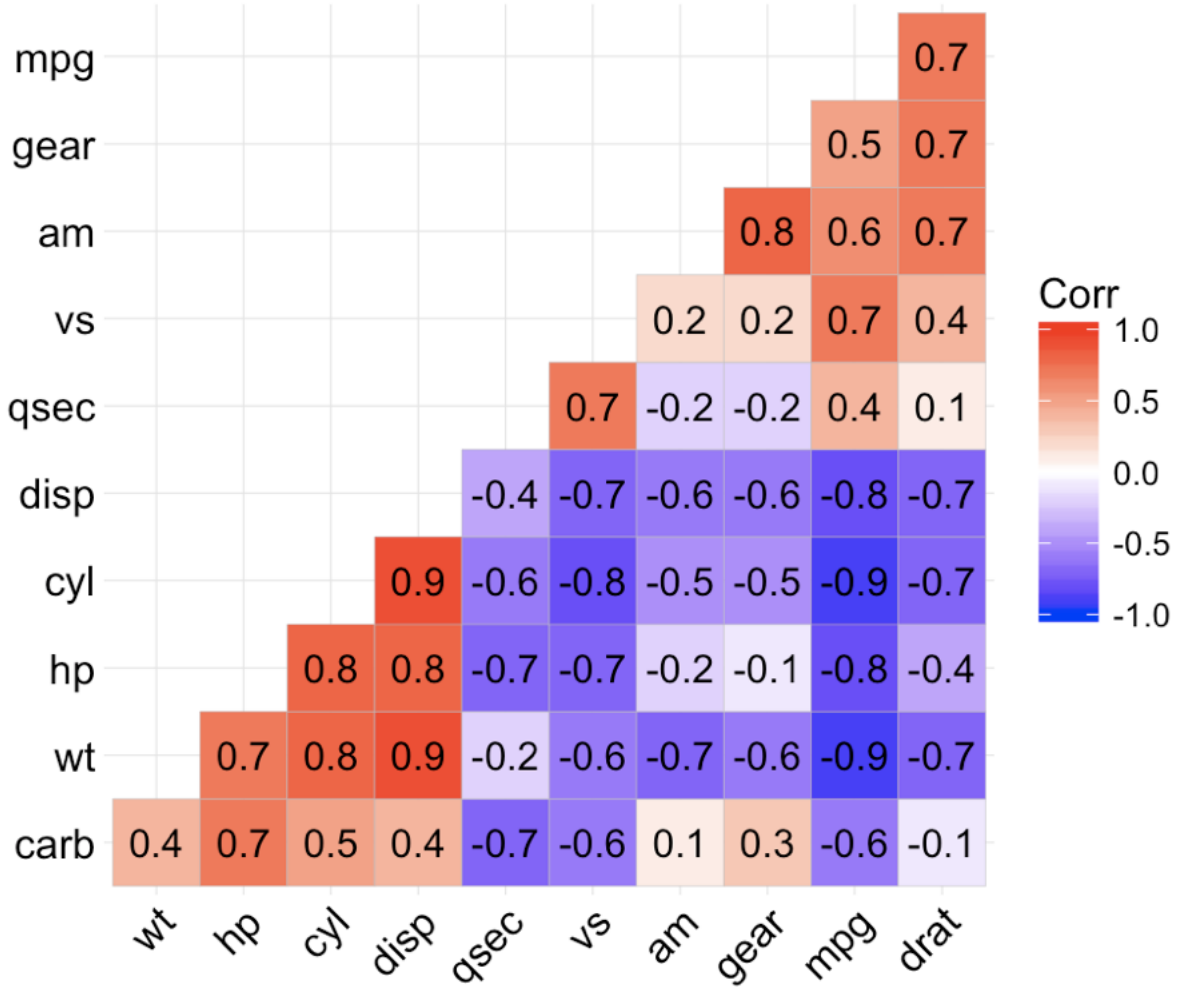


Figure 1: Example correlation plot.

instead have several binary variables corresponding to each racial category (e.g. 1 for Black, 0 otherwise; 1 for Hispanic, 0 otherwise, etc.) which can quickly become unwieldy. Try it both ways if you are interested in seeing the difference (though you'll need to turn the `race` variable to a factor).

We also need to adjust the vote choice variable. Currently, `presvote16` codes vote choice for any party in the 2016 election (Dem, Rep, Green, Libertarian) as well as votes for others. This, too, would become unwieldy in a regression. To simplify, we will turn this into an indicator variable for whether or not the respondent voted for the Republican (Donald J. Trump), called `rep_vote`. To be clear, `rep_vote` should be 1 if the respondent voted for Trump, 0 if they voted for someone else, and NA if they did not vote.

First, create the `white` variable from the `race` variable, as well as `rep_vote` from `presvote16`. Then, fit the linear model described above. Comment on what you see. Is this in line with what we would expect based on social dominance theory? Interpret your results and comment on what you find, especially as it relates to social dominance theory. Note that you can explore other model specifications in the next question.

```
sdo_new <- sdo %>%
  mutate(.keep="all", white=ifelse(race == 1, 1, 0), rep_vote=ifelse(presvote16=="Trump", 1, 0))
sdo_new
```

```
## # A tibble: 3,144 x 17
##     sdo5 female   age  educ  race favor_trump favor_blm favor_metoo
##    <dbl>  <dbl> <dbl> <dbl> <dbl>       <dbl>     <dbl>       <dbl>
##  1  3         1    75     2     1           4         1           1
##  2  2         0    73     5     1           4         1           2
##  3  1         0    69     2     1           1         4           4
##  4  2.5       0    79     6     1           4         1           2
##  5  1         0    64     5     1           1         4           4
##  6  1         0    60     3     7           1         4           4
##  7  1         1    49     6     2           1         4           4
##  8  3.5       0    66     5     1           4         1           1
##  9  3.5       0    76     6     1           4         1           1
## 10  1.75      0    38     6     1           1         4           4
## # ... with 3,134 more rows, and 9 more variables: american_customs <dbl>,
## #   race_ident <dbl>, pid3 <dbl>, ideo5 <dbl>,
## #   fear_of_demographic_change <dbl>, confederate_flag <chr>, presvote16 <chr>,
## #   white <dbl>, rep_vote <dbl>
```

```
fit <- lm(rep_vote ~ sdo5 + female + white + educ + age + pid3 + ideo5, data=sdo_new)
summary(fit)
```

```
##
## Call:
## lm(formula = rep_vote ~ sdo5 + female + white + educ + age +
##     pid3 + ideo5, data = sdo_new)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.05791 -0.14496  0.02687  0.15996  1.09151
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.6892609  0.0343974 -20.038  < 2e-16 ***
## sdo5         0.0935332  0.0074021  12.636  < 2e-16 ***
```

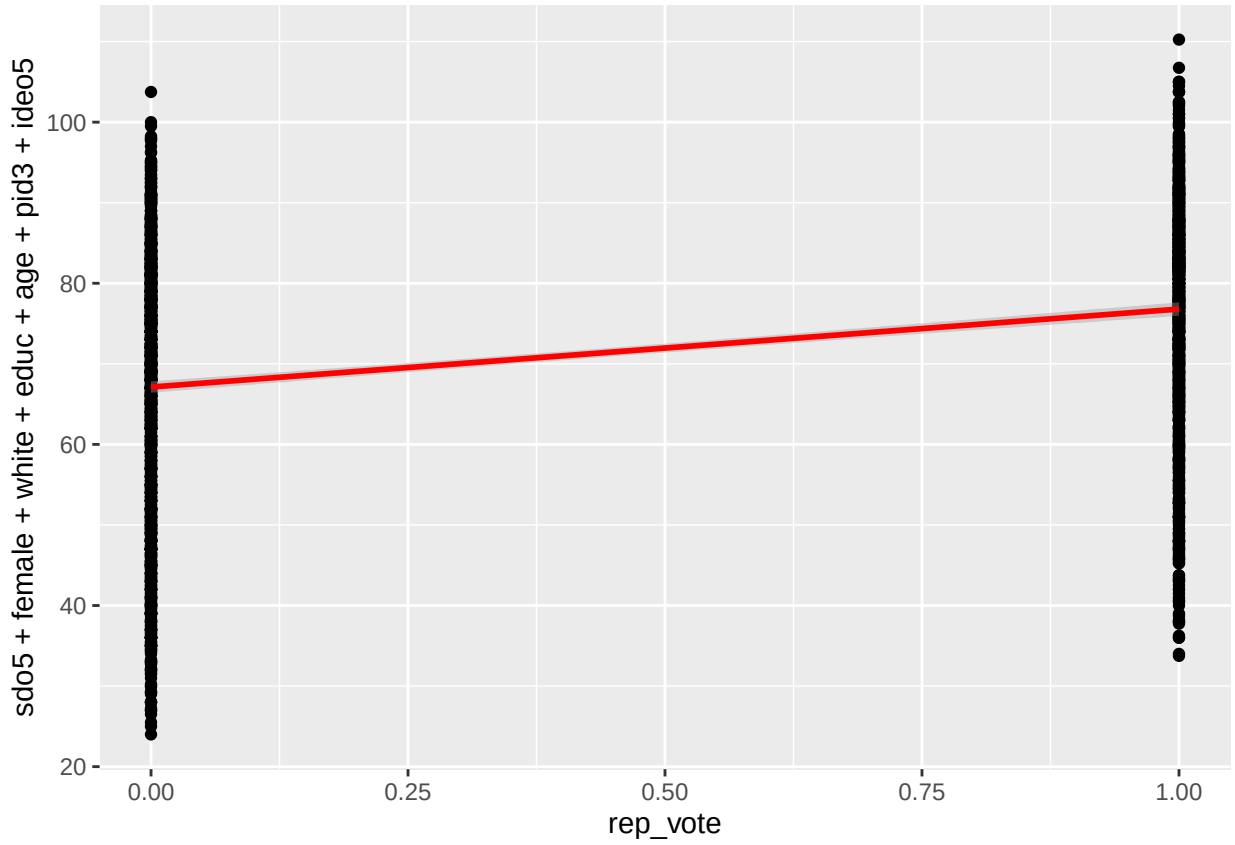
```
## female      -0.0035856  0.0115590  -0.310  0.75643
## white        0.0440001  0.0152242   2.890  0.00388 **
## educ        -0.0111305  0.0042179  -2.639  0.00836 **
## age          0.0020776  0.0003948   5.262 1.53e-07 ***
## pid3         0.2484180  0.0101457  24.485  < 2e-16 ***
## ideo5        0.1087947  0.0068317  15.925  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.297 on 2876 degrees of freedom
##   (260 observations deleted due to missingness)
## Multiple R-squared:  0.6394, Adjusted R-squared:  0.6385
## F-statistic: 728.4 on 7 and 2876 DF,  p-value: < 2.2e-16
```

```
ggplot(sdo_new, aes(x = rep_vote, y = sdo5 + female + white + educ + age + pid3 + ideo5)) +
  geom_point() +
  stat_smooth(method = "lm", col = "red")
```

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

```
## Warning: Removed 260 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 260 rows containing missing values (geom_point).
```



Question 7

Lastly, just explore the data! This question is open-ended, but make sure you have a theoretical expectation in mind for any relationships between variables you want to explore, and include them in your answer.

```
sdo2 <- sdo %>% mutate(confederate_flag = case_when(
  confederate_flag == "Heritage" ~ 1,
  confederate_flag == "Slavery" ~ 0,
  T ~ NA_real_
))
m1 <- lm(confederate_flag ~ fear_of_demographic_change, data=sdo2)

m2 <- lm(confederate_flag ~ fear_of_demographic_change + favor_metoo + age + sdo5, data=sdo2)

stargazer(m1, m2, header=FALSE, type="text")
```

```
##
## ================================================================================
##                                          Dependent variable:
##                                 ---------------------------------------------
##                                              confederate_flag
##                                        (1)                        (2)
## --------------------------------------------------------------------------------
## fear_of_demographic_change          1.131***                   0.465***
##                                      (0.026)                    (0.031)
##
## favor_metoo                                                    -0.188***
##                                                                 (0.007)
##
## age                                                            0.002***
##                                                                (0.0004)
##
## sdo5                                                           0.084***
##                                                                 (0.009)
##
## Constant                            0.053***                   0.524***
##                                      (0.013)                    (0.042)
##
## --------------------------------------------------------------------------------
## Observations                          2,908                      2,698
## R2                                    0.396                      0.582
## Adjusted R2                           0.396                      0.581
## Residual Std. Error             0.388 (df = 2906)          0.324 (df = 2693)
## F Statistic                1,903.738*** (df = 1; 2906) 936.437*** (df = 4; 2693)
## ================================================================================
## Note:                                              *p<0.1; **p<0.05; ***p<0.01
```