Data Exploration: Contextual Influences

Yanxi Fang

November 11, 2021

Developing a Research Question about Contextual Influences

Data Details:

- File Name: `vars_data.xlsx`

- Source: This file shows what variables are covered in each wave of the Nationscape Data Set (Tausanovitch and Vavreck 2020). We will be using data from the survey itself in other parts of the exercise, but which specific files and variables will be up to you! Therefore, we don't present them in depth here.

| Variable Name	Variable Description
Date	The date of the wave of the Nationscape survey
response_id	This and all other variables are the names of variables in the Nationscape data; the cells are 1 if that variable was included in that week's survey and 0 otherwise

```r
#Load the data summarizing variable availability
NationscapeVars_1 <- read_xlsx('vars_data.xlsx',sheet = 1)
NationscapeVars_2 <- read_xlsx('vars_data.xlsx',sheet = 2)
```

Now let's get the data from two sheets into one data set.

```r
NationscapeVars <- full_join(NationscapeVars_1,NationscapeVars_2) %>%
  replace(is.na(.),0)
```

Question 1

An event that happened in 2019 was a Boston federal judge's decision to uphold Harvard University's affirmative action admissions practices; this decision was released on October 1, 2019. As shown below, there is a spike in interest around the beginning of October 2019.
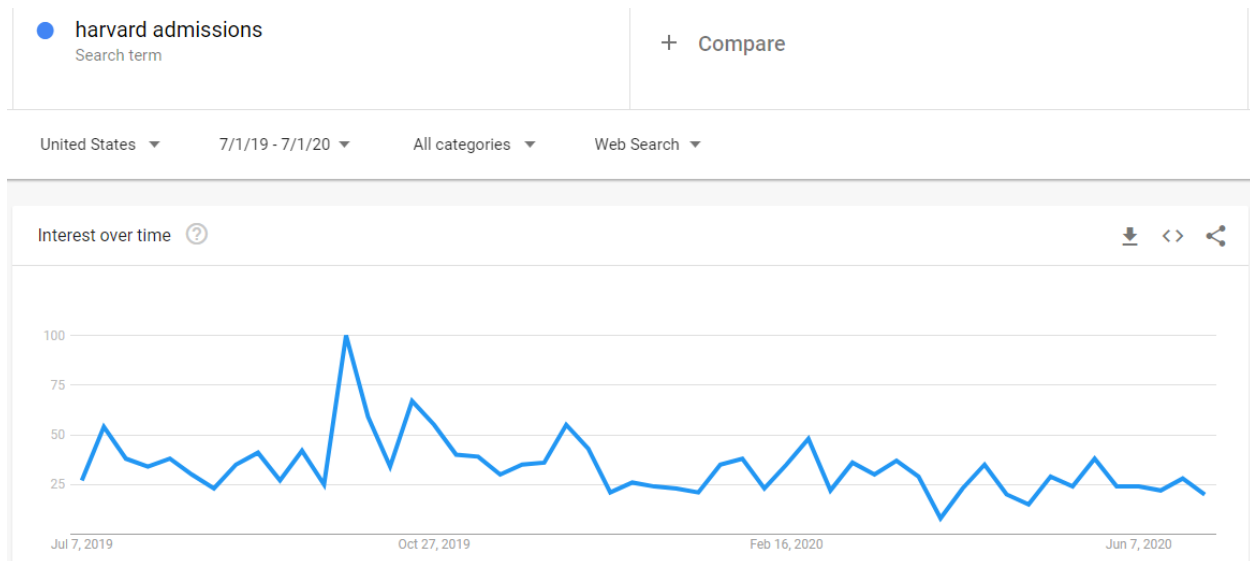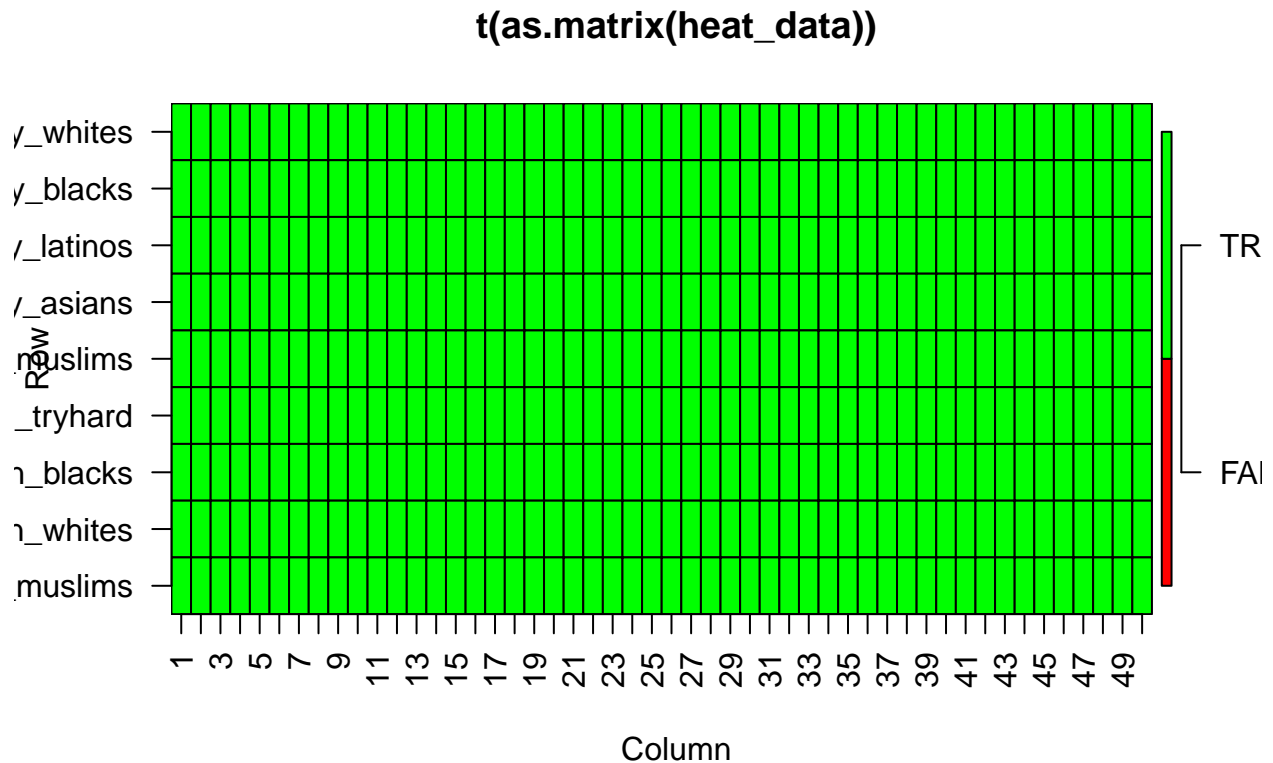
Question 2

There are three proxies related to Harvard admissions in the data: people's perceptions about discrimination faced by various racial groups in the US (three variables: Black, white, Muslim), general racial group favorability (five variables: Black, Latinx, white, Asian, Muslim), and the racial attitude statement comparing Black Americans to other minorities that overcame prejudice and worked their way up (`tryhard`). All of these might be affected by the decision because the lawsuit was inherently about racial preferences.

The availability of these data is checked using the heatmap. Although the variable names are not legible, the fact that the heatmap is entirely green indicates that all of the selected proxy variables are asked every week, so there is no concern about missing data.

```
heat_data <- NationscapeVars %>% mutate(across(.cols = everything(), as.logical)) %>%
  select(c(group_favorability_whites, group_favorability_blacks,
          group_favorability_latinos, group_favorability_asians,
          group_favorability_muslims, racial_attitudes_tryhard,
          discrimination_blacks, discrimination_whites, discrimination_muslims))

plot(t(as.matrix(heat_data)), col = c('red','green'), las = 2, par(cex = 0.5))
```

t(as.matrix(heat_data))

Question 3

My broader motivation here would be to try to understand whether Americans view the Harvard admissions case as one that is strongly tied to racial issues, as we might reasonably (or at least intuitively) expect. Thus, in this specific context, my research question is: "Was the October 2019 judicial ruling on the Harvard admissions case associated with changes in Americans' racial attitudes through priming?"

My hypothesis is that the Harvard admissions ruling was associated with changes in Americans' racial attitudes. This is an example of how current events can lead to increased awareness of certain issues (i.e. increase a topic's salience in one's associative network), thus causing the person to come up with a response that is more accurately reflective of their true beliefs (although we cannot directly observe people's true beliefs).

Question 4

There were two things that I took away from the group meeting.

First, there is a concern about how I would use the data to measure greater salience about racial attitudes. Specifically, increased salience does not translate neatly into the data: assuming that people (on average) change their responses due to increased salience, the change would not be in the same direction for everyone (i.e. some people see other races less favorably while others see other races more favorably), such that no change in the average attitudes might actually be reflective of greater salience. Upon hearing this, I thought that I can bypass this issue by adding the respondent's race as an explanatory variable, since one might assume that the direction in which attitudes shift might be consistent within racial groups (however, substantively, it can be problematic to assume that racial groups are monoliths).

The second piece of feedback was about my hypothesis. Instead of salience, it was suggested that I could alternatively explain any change in the outcome variables (the racial attitudes) using the emotion channel: the upholding of affirmative action policies may have led to a sense of continued loss for the racial groups that do not directly benefit, potentially leading to less tolerant attitudes toward the racial groups that are perceived to benefit. I think this is a very valid point; however, I am not seeing how I might be able to explain changes in attitudes among members of racial groups that are perceived to benefit, since the ruling merely reaffirmed an existing policy and therefore might not have generated much by way of emotional response.

Question 5

I found three articles related to the issue of racial responses to affirmative action; these are listed below.

Fine, Terri Susan. "The Impact of Issue Framing on Public Opinion: Toward Affirmative Action Programs." The Social Science Journal, Volume 29, Issue 3 (1992), pp. 323-334. https://doi.org/10.1016/0362-3319(92)90025-D.

Lausanne Renfro, C., et al. "The Role of Threat in Attitudes Toward Affirmative Action and Its Beneficiaries." Journal of Applied Psychology, Volume 36, Issue 1 (2006), pp. 41-74. https://doi.org/10.1111/j.0021-9029.2006.00003.x.

Richardson, John D. "Switching Social Identities: The Influence of Editorial Framing on Reader Attitudes Toward Affirmative Action and African Americans." Communication Research, Volume 32, Issue 4 (2005), pp. 503-528. https://doi.org/10.1177%2F0093650205277321.

Question 6

Below, I read in the data from five weeks' worth of surveys, making sure to recode the NAs. I then created a binary indicator to separate surveys that were administered before the court ruling on October 1, 2019, and surveys administered after that date. Subsequently, I ran a difference-in-means analysis on four variables: `racial_attitudes_tryhard`, `discrimination_blacks`, `discrimination_whites`, and `group_favorability_blacks`. These four variables were chosen from the list that I had identified earlier for various substantive reasons. For instance, the `tryhard` variable can intuitively be linked to affirmative action because affirmative action policies tend to benefit students from minority groups that do not do so well on standardized metrics of performance. Similarly, `discrimination_blacks` and `discrimination_whites` are on both sides of the same coin: affirmative action policies may be seen as benefitting Black Americans, helping lower the overall level of discrimination, but may also be seen as hurting white Americans.

The results are printed below. Three of the four variables did not even have statistically significant differences between the means; only `discrimination_blacks` did, and even then, according to Cohen's D, the size of the impact is considered negligible. Since the Nationscape survey is not a panel survey (i.e. it doesn't ask the same groups of people every week), it is possible that the change is simply due to a shift in the specific groups of people being sampled from week to week, in which case, there wouldn't be any evidence to support my hypothesis using this test. In addition, if the negative difference is to be taken at face value, it would mean that after the ruling, Americans thought that there was more discrimination against Black Americans (since lower values of the variable correspond to higher levels of discrimination), which would be the opposite of what I might expect based on the salience theory (i.e. since policies like affirmative action are in place to account for other types of discrimination, reminding Americans about affirmative action might make them feel that Black Americans suffer from less discrimination than what they might have felt pre-ruling).

```
# load the weekly survey files
Sep19 <- read_dta('ns20190919.dta') %>%
  remove_all_labels()
Sep26 <- read_dta('ns20190926.dta')%>%
  remove_all_labels()
```

```r
Oct03 <- read_dta('ns20191003.dta') %>%
  remove_all_labels()
Oct10 <- read_dta('ns20191010.dta')%>%
  remove_all_labels()
Oct17 <- read_dta('ns20191017.dta')%>%
  remove_all_labels()

# join them all together
Fall2019 <- full_join(Sep19,Sep26) %>%
  full_join(., Oct03) %>%
  full_join(., Oct10) %>%
  full_join(., Oct17)

# recode NAs
Fall2019 <- Fall2019 %>%
  mutate(across(.cols = everything(), ~na_if(., 999))) %>%
  mutate(across(.cols = everything(), ~na_if(., 888)))

# create an indicator variable for surveys administered after the court ruling
Fall2019 <- Fall2019 %>%
  mutate(treated = if_else(start_date > as.Date('2019-10-01'), TRUE, FALSE))

# difference in means for `tryhard`
difference_in_means(racial_attitudes_tryhard ~ treated, data = Fall2019 %>%
                    filter(as.Date('2019-09-26') < start_date) %>%
                    filter(start_date < as.Date('2019-10-03')))
```

```
## Design:  Standard
##          Estimate Std. Error  t value  Pr(>|t|)    CI Lower  CI Upper        DF
## treated 0.04979783 0.03994121 1.246778 0.2125894 -0.02852142 0.1281171 2638.734
```

```r
# effect size with Cohen's D
cohen.d(racial_attitudes_tryhard ~ treated, data = Fall2019 %>%
        filter(as.Date('2019-09-26') < start_date) %>%
        filter(start_date < as.Date('2019-10-03')))
```

```
## Warning in cohen.d.formula(racial_attitudes_tryhard ~ treated, data = Fall2019
## %>% : Cohercing rhs of formula to factor
```

```
##
## Cohen's d
##
## d estimate: -0.03688844 (negligible)
## 95 percent confidence interval:
##       lower       upper
## -0.09390600  0.02012911
```

```r
# same analysis for `discrimination_blacks`
difference_in_means(discrimination_blacks ~ treated, data = Fall2019 %>%
                    filter(as.Date('2019-09-26') < start_date) %>%
                    filter(start_date < as.Date('2019-10-03')))
```

```
## Design:   Standard
##            Estimate Std. Error   t value    Pr(>|t|)   CI Lower    CI Upper
## treated -0.1261566 0.03463446 -3.642516 0.0002750025 -0.1940689 -0.05824422
##               DF
## treated 2733.92
```

```r
cohen.d(discrimination_blacks ~ treated, data = Fall2019 %>%
          filter(as.Date('2019-09-26') < start_date) %>%
          filter(start_date < as.Date('2019-10-03')))
```

```
## Warning in cohen.d.formula(discrimination_blacks ~ treated, data = Fall2019
## %>% : Cohercing rhs of formula to factor
```

```
##
## Cohen's d
##
## d estimate: 0.1057304 (negligible)
## 95 percent confidence interval:
##      lower       upper
## 0.04850073 0.16296008
```

```r
# same analysis for `discrimination_whites`
difference_in_means(discrimination_whites ~ treated, data = Fall2019 %>%
                      filter(as.Date('2019-09-26') < start_date) %>%
                      filter(start_date < as.Date('2019-10-03')))
```

```
## Design:   Standard
##            Estimate Std. Error  t value   Pr(>|t|)     CI Lower  CI Upper
## treated 0.07045946 0.03702203 1.903177 0.0571256 -0.002135491 0.1430544
##               DF
## treated 2653.707
```

```r
cohen.d(discrimination_whites ~ treated, data = Fall2019 %>%
          filter(as.Date('2019-09-26') < start_date) %>%
          filter(start_date < as.Date('2019-10-03')))
```

```
## Warning in cohen.d.formula(discrimination_whites ~ treated, data = Fall2019
## %>% : Cohercing rhs of formula to factor
```

```
##
## Cohen's d
##
## d estimate: -0.05605277 (negligible)
## 95 percent confidence interval:
##         lower          upper
## -0.113438925   0.001333377
```

```r
# same analysis for `group_favorability_blacks`
difference_in_means(group_favorability_blacks ~ treated, data = Fall2019 %>%
                      filter(as.Date('2019-09-26') < start_date) %>%
                      filter(start_date < as.Date('2019-10-03')))
```

```
## Design:  Standard
##          Estimate Std. Error  t value  Pr(>|t|)   CI Lower   CI Upper       DF
## treated 0.03707233  0.0262867 1.410308 0.1585827 -0.0144755 0.08862017 2323.844
```

```
cohen.d(group_favorability_blacks ~ treated, data = Fall2019 %>%
          filter(as.Date('2019-09-26') < start_date) %>%
          filter(start_date < as.Date('2019-10-03')))
```

```
## Warning in cohen.d.formula(group_favorability_blacks ~ treated, data = Fall2019
## %>% : Cohercing rhs of formula to factor
```

```
##
## Cohen's d
##
## d estimate: -0.04434301 (negligible)
## 95 percent confidence interval:
##      lower      upper
## -0.10579000  0.01710399
```

Question 7: DATA SCIENCE QUESTION

To extend upon my work from the previous question, I decided to add in confounding variables to the analysis. This decision was motivated by the feedback that I had received about salience not being reflected if different racial groups move in different directions on variables after hearing about the court ruling. It was also motivated by the fact that it is substantively unrealistic to expect all Americans to react the same way to the rulings, a detail that makes it important to control for differences among Americans.

To operationalize the analysis with the confounding variables, I ended up running a series of multiple linear regression models. The outcome variables were the four that I had analyzed in the previous question: `tryhard`, `discrimination_blacks`, `discrimination_whites`, and `group_favorability_blacks`. The confounding variables I chose were: a binary indicator of whether the respondent is Hispanic, a binary indicator of whether the respondent is white/non-white, education, `reparations` (a variable indicating the respondent's support for monetary compensation for the descendents of slaves), the `ideo5` ideology indicator, the party ID, as well as household income. These were generally chosen because they are demographic confounders, with the `reparations` variable chosen because substantively, one's support for reparations should have a high correlation with a belief that Black Americans face high levels of discrimination, which greatly influences the outcome variable even when holding everything else constant.

The results are printed in Table 2. As shown, while various confounders have statistically significant coefficients, meaning that their impact on the outcome variable is non-zero, the same cannot be said for our main variable of interest, `treated`, across all four models. In other words, this model suggests that the treatment – the contextual exposure to the October 1, 2019 court ruling that upheld Harvard's affirmative action admissions policies – did not have a substantial impact on how Americans felt about Black Americans and the discrimination they face, thus suggesting that my hypothesis is not supported by the data.

```
# create new variables for hispanic and non-white
Fall2019 <- Fall2019 %>%
  mutate(binary_hisp = ifelse(hispanic == 1, 0, 1)) %>%
  mutate(binary_white = ifelse(race_ethnicity == 1, 1, 0))

# run linear regression on the four variables, with various control variables
q7a <- lm(racial_attitudes_tryhard ~ treated + binary_hisp + binary_white +
                              education + reparations + ideo5 +
```

```
                                  as.factor(pid3) + household_income,
         data = Fall2019)
q7b <- lm(discrimination_blacks ~ treated + binary_hisp + binary_white +
                                  education + reparations + ideo5 +
                                  as.factor(pid3) + household_income,
         data = Fall2019)
q7c <- lm(discrimination_whites ~ treated + binary_hisp + binary_white +
                                  education + reparations + ideo5 +
                                  as.factor(pid3) + household_income,
         data = Fall2019)
q7d <- lm(group_favorability_blacks ~ treated + binary_hisp + binary_white +
                                  education + reparations + ideo5 +
                                  as.factor(pid3) + household_income,
         data = Fall2019)
```

```
# print results
stargazer(q7a, q7b, q7c, q7d, font.size = "tiny",
         column.sep.width = "0pt", header = FALSE,
         title = "Impacts of the October 2019 Harvard Admissions Ruling on American Racial Attitudes")
```

Table 2: Impacts of the October 2019 Harvard Admissions Ruling on American Racial Attitudes

	Dependent variable:			
	racial_attitudes_tryhard	discrimination_blacks	discrimination_whites	group_favorability_blacks
	(1)	(2)	(3)	(4)
treated	0.029	−0.019	0.026	0.019
	(0.031)	(0.025)	(0.031)	(0.022)
binary_hisp	−0.147***	−0.035	−0.021	0.071**
	(0.044)	(0.036)	(0.044)	(0.031)
binary_white	−0.239***	0.262***	−0.400***	0.042
	(0.039)	(0.032)	(0.039)	(0.028)
education	0.037***	−0.007	0.023***	0.004
	(0.008)	(0.006)	(0.008)	(0.005)
reparations	−0.666***	0.752***	−0.122***	0.276***
	(0.036)	(0.030)	(0.036)	(0.025)
ideo5	−0.211***	0.199***	−0.096***	0.031***
	(0.015)	(0.013)	(0.015)	(0.011)
as.factor(pid3)2	−0.478***	0.540***	−0.547***	0.173***
	(0.042)	(0.035)	(0.042)	(0.030)
as.factor(pid3)3	−0.136***	0.265***	−0.202***	0.111***
	(0.039)	(0.032)	(0.039)	(0.028)
as.factor(pid3)4	0.070	0.316***	−0.444***	0.204***
	(0.085)	(0.071)	(0.085)	(0.061)
household_income	−0.002	0.0001	0.001	−0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	4.349***	0.089	4.267***	1.122***
	(0.081)	(0.067)	(0.081)	(0.057)
Observations	7,153	7,103	7,087	6,530
R^2	0.210	0.313	0.104	0.054
Adjusted R^2	0.209	0.312	0.103	0.053
Residual Std. Error	1.226 (df = 7142)	1.007 (df = 7092)	1.219 (df = 7076)	0.822 (df = 6519)
F Statistic	189.423*** (df = 10; 7142)	322.875*** (df = 10; 7092)	82.253*** (df = 10; 7076)	37.245*** (df = 10; 6519)

Note: *p<0.1; **p<0.05; ***p<0.01