Data Exploration: Symbolic Politics

Paddy Adams

October 21, 2021

```r
load('RN_2001_data.RData')
```

```r
protest_df_bydate <- protest_df %>%  mutate(before = ifelse(day < as.Date("2020-05-28"), 1,0))
```

For this weeks data we explored Reny and Newman's (2021) finding that in the wake of George Floyd's killing the resulting spread of protests influenced opinions about the level of discrimination faced by black Americans as well as the group favourability of Police. The data includes many other points of interest and variables but these aforementioned two are the main focus. To begin with it is worth seeing if the average for the outcome variables both before and after the 28th of May is statistically significant: for the variable 'group_favorability_the_police' the resulting t-test p-value was 0.02366, which we can conclude makes it a statistically significant change in mean. For 'discrimination_blacks' the resulting p-value was again significant at 0.01174.

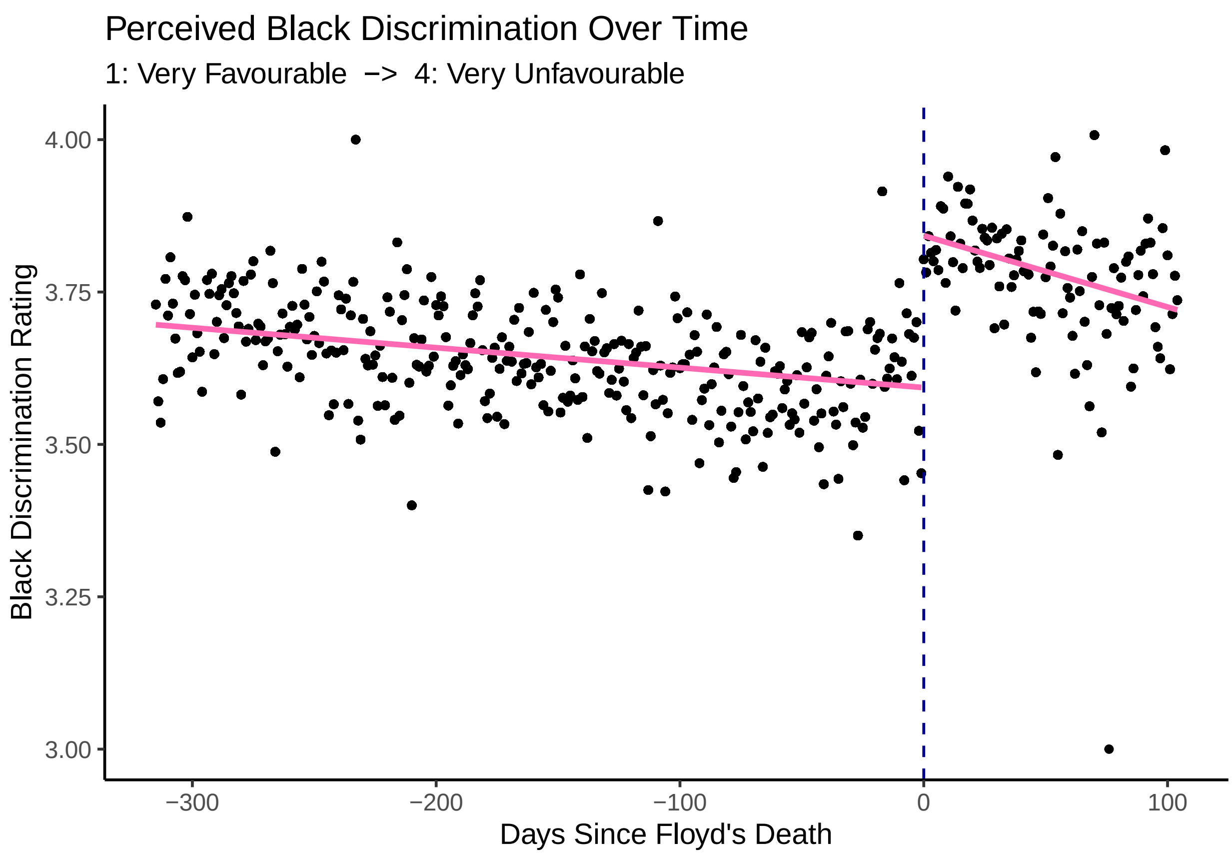
```r
# was confused by t.test results in chunks below (they don't...add up) so did some playing around

ttest <- protest_df_bydate %>%
  group_by(before) %>%
  select(11, 12, 19, 23) %>%
  summarise_all(mean, na.rm = TRUE)

a <- ttest %>%
  select(2) %>%
  t.test()

b <- ttest %>%
  select(3) %>%
  t.test()



# protest_df_bydate$group_favorability_the_police <- as.numeric(protest_df_bydate$group_favorability_th
# protest_df_bydate$before <- as.character(protest_df_bydate$before)

sd_mean_etc <- protest_df_bydate %>%
  select(11, 23) %>%
  na.omit() %>%
  group_by(before) %>%
  summarise(
    count = n(),
    mean = mean(group_favorability_the_police),
    sd = sd(group_favorability_the_police))
```

```r
pre <- protest_df_bydate %>%
  select(11, 23) %>%
  na.omit() %>%
  filter(before == 1) %>%
  head(86849)

post <- protest_df_bydate %>%
  select(11, 23) %>%
  na.omit() %>%
  filter(before == 0) %>%
  head(86849)

g <- post %>% full_join(pre)

v <- t.test(group_favorability_the_police ~ before, data = g, paired = TRUE)
```

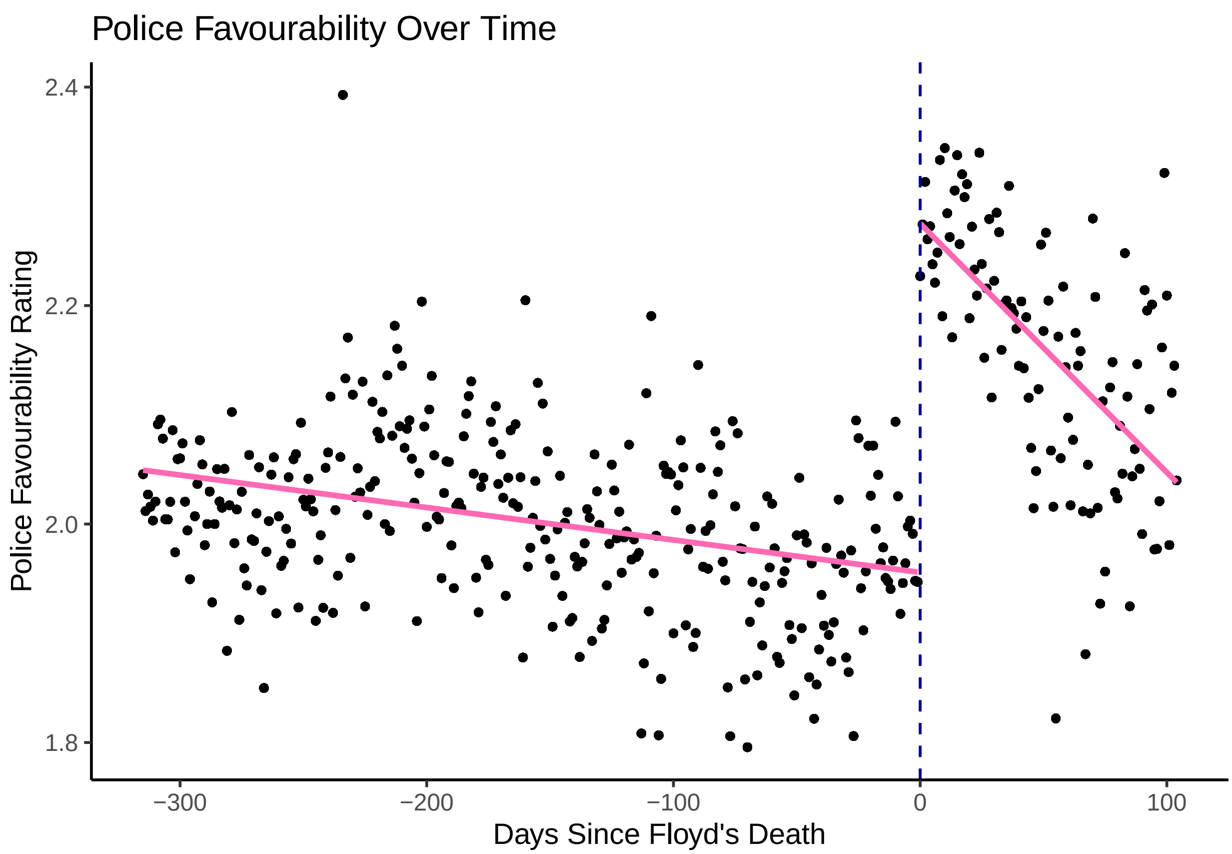
In order to better visualize this here are some plots of the two outcome variables we are looking at:

```r
avg_by_day <- protest_df_bydate %>%
  select(1, 2, 3, 11, 12, 23) %>%
  group_by(day_running) %>%
  mutate(avg_fav_pol = mean(group_favorability_the_police, na.rm = T)) %>%
  mutate(avg_disc_bl = mean(discrimination_blacks, na.rm = T))
```

```r
avg_by_day %>%
  group_by(day_running) %>%
  ggplot(aes(x = day_running, y = avg_disc_bl, fill = as.factor(before))) +
  geom_point(colour = "black", size = 1) +
  geom_smooth(method="lm", se=FALSE, colour = "hotpink") +
  geom_vline(xintercept = 0, colour = 'darkblue', linetype = "dashed") +
  theme_classic() +
  theme(legend.position = 'none') +
  labs(title = "Perceived Black Discrimination Over Time",
       subtitle = "1: Very Favourable  ->  4: Very Unfavourable",
       x = "Days Since Floyd's Death",
       y = "Black Discrimination Rating")
```




```
avg_by_day %>%
  group_by(day_running) %>%
  filter(avg_fav_pol > 1.1) %>%
  ggplot(aes(x = day_running, y = avg_fav_pol, fill = as.factor(before))) +
  geom_point(colour = "black", size = 1) +
  geom_smooth(method="lm", se=FALSE, colour = "hotpink") +
  geom_vline(xintercept = 0, colour = 'darkblue', linetype = "dashed") +
  theme_classic() +
  theme(legend.position = 'none') +
  labs(title = "Police Favourability Over Time",
       x = "Days Since Floyd's Death",
       y = "Police Favourability Rating")
```



Of course, there are many potentially compounding factors here that may change the way people respond so it is worth looking at how attitudes towards other groups changed in the same period.

```
# potentially difference in n values gives skewed results

m <- protest_df_bydate %>%
  group_by(before) %>%
  select(11, 12, 17, 18, 19, 20, 23) %>%
  summarise_all(mean, na.rm = TRUE) %>%
  map_df(~ tidy(t.test(.)), .id = 'col')
```

```
# similar n values

dap <- protest_df_bydate %>%
  group_by(before) %>%
  select(11, 12, 17, 18, 19, 20, 23) %>%
  na.omit()

test <- dap %>%
  summarise_all(mean) %>%
  map_df(~ tidy(t.test(.)), .id = 'column') %>%
  select(column, p.value) %>%
```

```r
  tibble()

dap %>%
  summarise_all(mean) %>%
  mutate_all(funs(lag(.) - .)) %>%
  tidy() %>%
  select(column, mean) %>%
  tibble() %>%
  full_join(test) %>%
  rename(change_in_mean = mean,
         t.test_p.value = p.value) %>%
  tail(-1) %>%
  gt() %>%
  fmt_number(columns = 2:3, decimals = 4)
```

column	change_in_mean	t.test_p.value
group_favorability_the_police	0.1501	0.0238
discrimination_blacks	0.1372	0.0117
group_favorability_jews	−0.0168	0.0031
group_favorability_whites	−0.0461	0.0046
group_favorability_evangelicals	−0.0255	0.0035
group_favorability_socialists	−0.0108	0.0013

```r
# note for favourability it goes 1: very fav to 4: very unfav
# and for disc_blacks - 1: none to 5: a lot
```

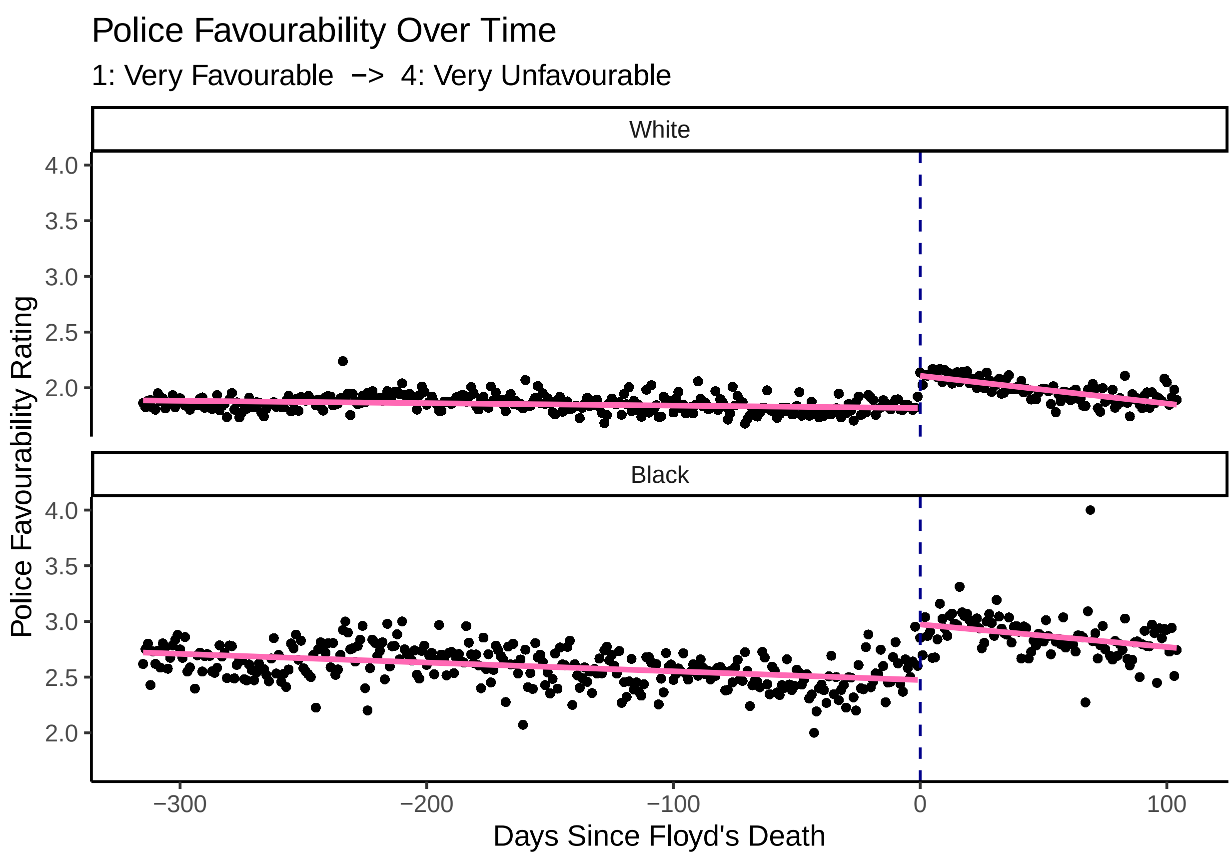
As we can see this produces some interesting results - while the two variables we have already seen both increase on average by a statistically significant amount (an increase means a more negative response), the other four favorability variables all show a small but statistically significant increase in their respective group favourability (decrease in variable value). This may be because people became more aware of their attitudes towards all groups and tried to behave more morally righteous - that is to say responding more how they would knowing maybe others were judging them in a similar way to what we explored in week one with the work Gerber, Green, and Larimer as part of their experiment on Michigan voters in the context of the 2006 primary election. Or it may be that these changes are as an honest result of the decrease in police favourability, as if this value decreases then the relative view of other groups is going to increase regardless of them having done anything actionable (such that in this context the score increases).

It should be noted that while I was looking at these results I couldn't help but think the figures didn't quite 'add up' (more 'significant' variables with a smaller change in mean). So after a lot of time spent on a large tangent (some of which can be seen in my code) I concluded that I had probably carried out the t-tests wrong. Nonetheless I have left this in if they were done correctly it is noteworthy.

Due to the nature of the protest's subject matter, it was particularly interesting to see how the respondents opinons changed when grouping by race. Again, this is easier to show graphically:

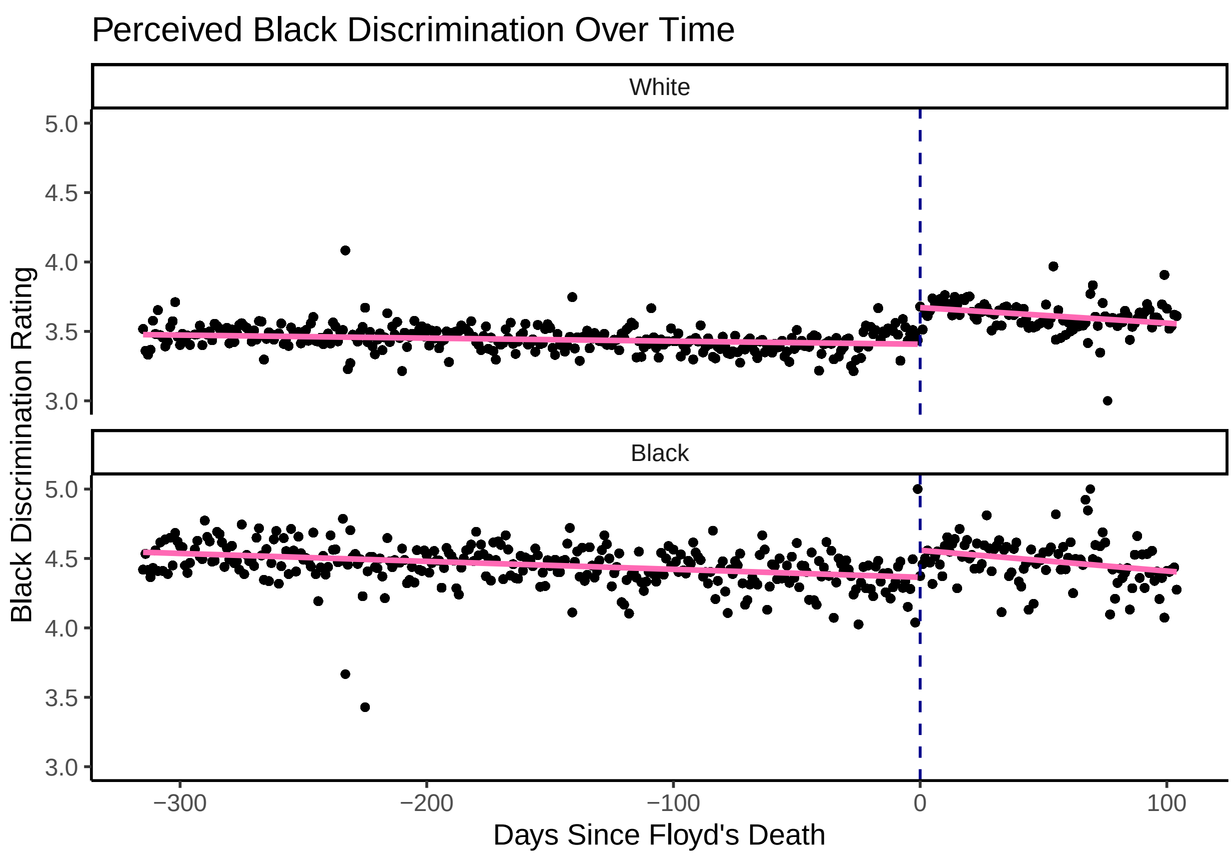
```r
avg_by_day_race <- protest_df_bydate %>%
  select(1, 2, 3, 11, 12, 23) %>%
  filter(race_ethnicity == 1 | race_ethnicity == 2) %>%
  filter(hispanic == 1) %>%
  select(-hispanic) %>%
  group_by(race_ethnicity, day_running) %>%
  mutate(avg_fav_pol = mean(group_favorability_the_police, na.rm = T)) %>%
```

```
  mutate(avg_disc_bl = mean(discrimination_blacks, na.rm = T))

avg_by_day_race$race_ethnicity2 <- factor(avg_by_day_race$race_ethnicity, labels = c("White", "Black"))

avg_by_day_race %>%
  group_by(day_running) %>%
  filter(avg_fav_pol > 1.1) %>%
  ggplot(aes(x = day_running, y = avg_fav_pol, fill = as.factor(before))) +
  geom_point(colour = "black", size = 1) +
  geom_smooth(method="lm", se=FALSE, colour = "hotpink") +
  geom_vline(xintercept = 0, colour = 'darkblue', linetype = "dashed") +
  facet_wrap(~ race_ethnicity2, dir = "v") +
  theme_classic() +
  theme(legend.position = 'none') +
  labs(title = "Police Favourability Over Time",
       subtitle = "1: Very Favourable  ->  4: Very Unfavourable",
       x = "Days Since Floyd's Death",
       y = "Police Favourability Rating")
```




```
avg_by_day_race %>%
  group_by(day_running) %>%
  ggplot(aes(x = day_running, y = avg_disc_bl, fill = as.factor(before))) +
  geom_point(colour = "black", size = 1) +
  geom_smooth(method="lm", se=FALSE, colour = "hotpink") +
  geom_vline(xintercept = 0, colour = 'darkblue', linetype = "dashed") +
```

```
facet_wrap(~ race_ethnicity2, dir = "v") +
theme_classic() +
theme(legend.position = 'none') +
labs(title = "Perceived Black Discrimination Over Time",
     x = "Days Since Floyd's Death",
     y = "Black Discrimination Rating")
```



There is a large but understandable discrepancy in the averages before and after the incident, with 'white' respondents typically a whole number, lower (more favourable/less discrimination) than 'black' respondents. What is interesting is how these numbers changed after the 'day 0'. Police favourability decreased for black respondents to a larger degree, and didn't fall off as quickly - white respondents were back at pre-incident levels after 100 days where black ratings had only returned by a half. For neither of the two demographics did perceived black discrimination change dramatically, but this time it was the black respondents who returned to pre-incident levels faster. This is probably because they started so high to being with, while 'white' respondents had a genuine perception change.

One last thing to consider with these last graphics is non-response bias. This is where those who answer the surveys differ from those who don't over time. In this context it is relevant as when non-response is correlated with the outcomes being measured it is harder to infer from the data. To use the example given in the prompt: "For example after a series of damaging headlines supporters of a politician may be less willing to answer phone surveys about that politician. As a result we would potentially observe an exaggeration of the negative effects of the scandal on a politician's polled approval rating". As the killing of George Floyd sparked such wide spread racial tensions, it would be reasonable to assume those that do support the police may be non-respondents. However the responses for this survey stay very consistent, suggesting this isn't to large a factor.

```
response_bias <- protest_df_bydate %>%
  select(1, 2, 3, 11, 12, 23) %>%
  filter(race_ethnicity == 1 | race_ethnicity == 2) %>%
  filter(hispanic == 1) %>%
  select(-hispanic) %>%
  group_by(race_ethnicity, day_running, before) %>%
  count()

response_bias$race_ethnicity2 <- factor(response_bias$race_ethnicity, labels = c("White", "Black"))

ggplot(response_bias, aes(x = day_running, y = n, fill = as.factor(before))) +
  geom_point(colour = "black", size = 1) +
  geom_smooth(method="lm", se = FALSE, colour = "hotpink") +
  geom_vline(xintercept = 0, colour = 'darkblue', linetype = "dashed") +
  facet_wrap(~ race_ethnicity2, dir = "v") +
  theme_classic() +
  theme(legend.position = 'none')
```