Data Exploration: Symbolic Politics

Kai McNamee

October 21, 2021

In this Data Exploration assignment we will explore Reny and Newman's (2021) finding that opinions towards the police and about the level of discrimination faced by Black Americans were impacted by the spread of protests in the wake of the killing of George Floyd. You will recreate, present, and assess those claims as well as creating your own regression models to test which attitudes change and when.

If you have a question about any part of this assignment, please ask! Note that the actionable part of each question is **bolded**.

Opinion Mobilization: The George Floyd Protests

Data Details:

- File Name: `RN_2001_data.RData`
- Source: These data are from Reny and Newman (2021).

Variable Name	Variable Description
`race_ethnicity`	Race or ethnicity. Levels labelled in data: 1-White, 2-Black or AfAm, 3-American Indian or Alaskan Native, 4 through 14-Asian or Pacific Islander (details in labels), and 15-Some other race
`hispanic`	Of Hispanic, Latino, or Spanish origin. Levels labelled in data: 1-Not Hispanic, 2-15 Hispanic of various origins
`day_running`	Day relative to onset of George Floyd protests (day 0)
`age`	Respondent's age
`female`	Binary indicator variable: 1 if respondent female, 0 otherwise
`college`	Binary indicator variable: 1 if respondent attended college, 0 otherwise
`household_income`	Household pre-tax income ranging from 1 (less than $15,000) to 24 (more than $250,000). Details for other levels in labels.
`pid7`	Party identification on a seven point scale with strong, weak, lean: 1-Strong Democrat to 7-Strong Republican with 4-Independent.
`ideo5`	Ideological self placement: 1-Very liberal, 2-Liberal, 3-Moderate, 4-Conservative, 5-Very Conservative
`vote_clinton`	Indicator variable for whether the respondent said they voted for Clinton in the 2016 presidential election
`group_favorability_the_police`	Favorability towards the police: 1-Very favorable, 2-Somewhat favorable, 3-Somewhat unfavorable, 4-Very unfavorable

Variable Name	Variable Description
discrimination_blacks	Perceptions of the level of discrimination in US faced by Blacks: 1-None at all, 2-A little, 3-A moderate amount, 4-A lot, 5-A great deal
date	The date the respondent took the survey
group_fav_white_black	The difference in respondents favorability towards Blacks subtracted from their favorability towards whites (each on four point scale). Ranges from -3 to 3.
racial_attitudes_generations	Agreement with the statement that generations of slavery and discrimination have made it difficult for Blacks to work their way out of the lower class: 1-Strongly Agree to 5-Strongly Disagree
interest	Degree to which respondent claims to follow politics: 1-Most of the time, 2-Some of the time, 3-Only now and then, 4-Hardly at all
group_favorability_jews	Favorability towards Jews: 1-Very favorable, 2-Somewhat favorable, 3-Somewhat unfavorable, 4-Very unfavorable
group_favorability_whites	Favorability towards whites: 1-Very favorable, 2-Somewhat favorable, 3-Somewhat unfavorable, 4-Very unfavorable
group_favorability_evangelicals	Favorability towards evangelicals: 1-Very favorable, 2-Somewhat favorable, 3-Somewhat unfavorable, 4-Very unfavorable
group_favorability_socialists	Favorability towards socialists: 1-Very favorable, 2-Somewhat favorable, 3-Somewhat unfavorable, 4-Very unfavorable
protest	Indicator variable if survey respondent lived in area that would at any point have a BLM protest in the wake of the killing of George Floyd
n_protests	Number of eventual BLM protests in area where resident lived

```r
# load the data containing the tibble protest_df
load('RN_2001_data.RData')

#Note that the data is saved in the form of a tibble, a special table using the dplyr package that has 

head(protest_df$race_ethnicity)
```

```
## <labelled<double>[6]>: What is your race? Provided by LUCID.
## [1] 6 1 1 2 1 1
##
## Labels:
##  value                              label
##      1                              White
##      2          Black, or African American
##      3   American Indian or Alaska Native
##      4                Asian (Asian Indian)
##      5                    Asian (Chinese)
##      6                   Asian (Filipino)
##      7                   Asian (Japanese)
##      8                     Asian (Korean)
##      9                 Asian (Vietnamese)
##     10                      Asian (Other)
##     11 Pacific Islander (Native Hawaiian)
##     12       Pacific Islander (Guamanian)
##     13          Pacific Islander (Samoan)
```

```
##     14          Pacific Islander (Other)
##     15                   Some other race
##    777               Not asked in this wave
```

```
head(protest_df$household_income)
```

```
## <labelled<double>[6]>: What is your current annual household income before taxes? Provided by L...
## [1] 21  8  7  1 NA  1
##
## Labels:
##  value                  label
##      1      Less than $14,999
##      2     $15,000 to $19,999
##      3     $20,000 to $24,999
##      4     $25,000 to $29,999
##      5     $30,000 to $34,999
##      6     $35,000 to $39,999
##      7     $40,000 to $44,999
##      8     $45,000 to $49,999
##      9     $50,000 to $54,999
##     10     $55,000 to $59,999
##     11     $60,000 to $64,999
##     12     $65,000 to $69,999
##     13     $70,000 to $74,999
##     14     $75,000 to $79,999
##     15     $80,000 to $84,999
##     16     $85,000 to $89,999
##     17     $90,000 to $94,999
##     18     $95,000 to $99,999
##     19   $100,000 to $124,999
##     20   $125,000 to $149,999
##     21   $150,000 to $174,999
##     22   $175,000 to $199,999
##     23   $200,000 to $249,999
##     24     $250,000 and above
##    777 Not asked in this wave
```

Question 1

As usual it is important to first examine the structure of the data. What are the two main outcome variables of interest to Reny and Newman? How were they measured and how are they coded in the data? What was the treatment? **Take a look at the data and determine which are the two outcome variables of interest. Observe the scale of each.**

Reny and Newman use data from the Nationscap suvey (NS), a survey administered by the Democracy Fund and UCLA. The NS is a weekly survey with 6,250 respondents per week and an average of 900 respondents per day,

Reny and Newman measure 2 outcome variables:

1. group_favorability_the_police: survey response indicating how strongly an individual favors the police (1 is strongest favorability, 4 is least favorability)
2. discrimination_blacks: survey response indicating how much discrimination respondent perceives Black people experience (1 is least, 4 is most)

##Question 2

###Part a R has a special 'date' class for storing and manipulating dates as seen below. Date variables can conveniently be logically compared and arithmetically manipulated. Using the day variable find out how many days the dataset spans. **First check using the code below that the day variable is of the class 'date'. Next subtract the latest day in the sample from the first day to calculate the timespan covered by the dataset. Hint: functions like max() and min() work for date variables too!**

```
class(protest_df$day)
```

```
## [1] "Date"
```

```
protest_df %>%
  summarize(days = max(day) - min(day))
```

```
## # A tibble: 1 x 1
##   days
##   <drtn>
## 1 419 days
```

###Part b On what date is the treatment said to have occurred? **Find the date for which the day_running variable is 0. Then modify the code below to add a variable to each row for whether or not the observation was before or after treatment.**

```
# Change the object to be the date of the protest spread, remember to put it in
# quotes if you copy/paste!

# Day of treatment 2020-05-28

protest_df_bydate <- protest_df %>%
  mutate(treatment = ifelse(day_running < 0, 0, 1)) %>%
  drop_na()
```

Question 3

###Part a Compare the average for each outcome variable before and after the onset of the protests. Are the differences statistically significant? **Calculate the outcome variable means for before and after treatment. Conduct a test as to whether the differences in means are statistically significant. Hint: you can use either the t.test() function or difference_in_means() from the estimatr package**

```
protest_df_bydate %>%
  group_by(treatment) %>%
  summarize(police = mean(group_favorability_the_police),
            discrimination_mean = mean(discrimination_blacks), .groups = "drop")
```

```
## # A tibble: 2 x 3
##   treatment police discrimination_mean
##       <dbl>  <dbl>               <dbl>
## 1         0   1.89                3.64
## 2         1   2.03                3.77
```

```r
# Difference in means test for police favorability

t.test(protest_df_bydate$group_favorability_the_police[protest_df_bydate$treatment > 0], protest_df_byda
```

```
##
##  Welch Two Sample t-test
##
## data:  protest_df_bydate$group_favorability_the_police[protest_df_bydate$treatment > 0] and protest_
## t = 25.127, df = 99274, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1336424 0.1562555
## sample estimates:
## mean of x mean of y
##  2.031622  1.886673
```

```r
# Difference in means test for perceived discrimination

t.test(protest_df_bydate$discrimination_blacks[protest_df_bydate$treatment > 0], protest_df_bydate$disc
```

```
##
##  Welch Two Sample t-test
##
## data:  protest_df_bydate$discrimination_blacks[protest_df_bydate$treatment > 0] and protest_df_bydate
## t = 19.47, df = 104216, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1206434 0.1476516
## sample estimates:
## mean of x mean of y
##  3.770849  3.636702
```

###Part b It might be that the period before and after the treatment was different in ways in addition to the onset of the protests. Use the same procedure as above to check for differences between two means of a survey response measuring favorability towards a group besides the police. **Calculate the means from before and after the treatment and conduct a test of statistical significance of the difference for another measure of group favorability that was recorded in the survey (e.g. evangelicals, Jews, socialists, or whites). Is there also a substantive or statistically significant difference on that variable? Should that change our confidence in attributing the opinion changes found in part a to the George Floyd protests?**

```r
protest_df_bydate %>%
  group_by(treatment) %>%
  summarize(police = mean(group_favorability_the_police),
            discrimination_mean = mean(discrimination_blacks),
            evangelicals_mean = mean(group_favorability_evangelicals),
            socialists_mean = mean(group_favorability_socialists), .groups = "drop")
```

```
## # A tibble: 2 x 5
##   treatment police discrimination_mean evangelicals_mean socialists_mean
##       <dbl>  <dbl>               <dbl>             <dbl>           <dbl>
## 1         0   1.89                3.64              2.29            2.64
## 2         1   2.03                3.77              2.27            2.63
```

```r
# conduct t test comparing group_favorability_socialists when treatment > 0 and
# treatment == 0

mean(protest_df_bydate$group_favorability_socialists[protest_df_bydate$treatment > 0])
```

```
## [1] 2.634548
```

```r
mean(protest_df_bydate$group_favorability_socialists[protest_df_bydate$treatment == 0])
```

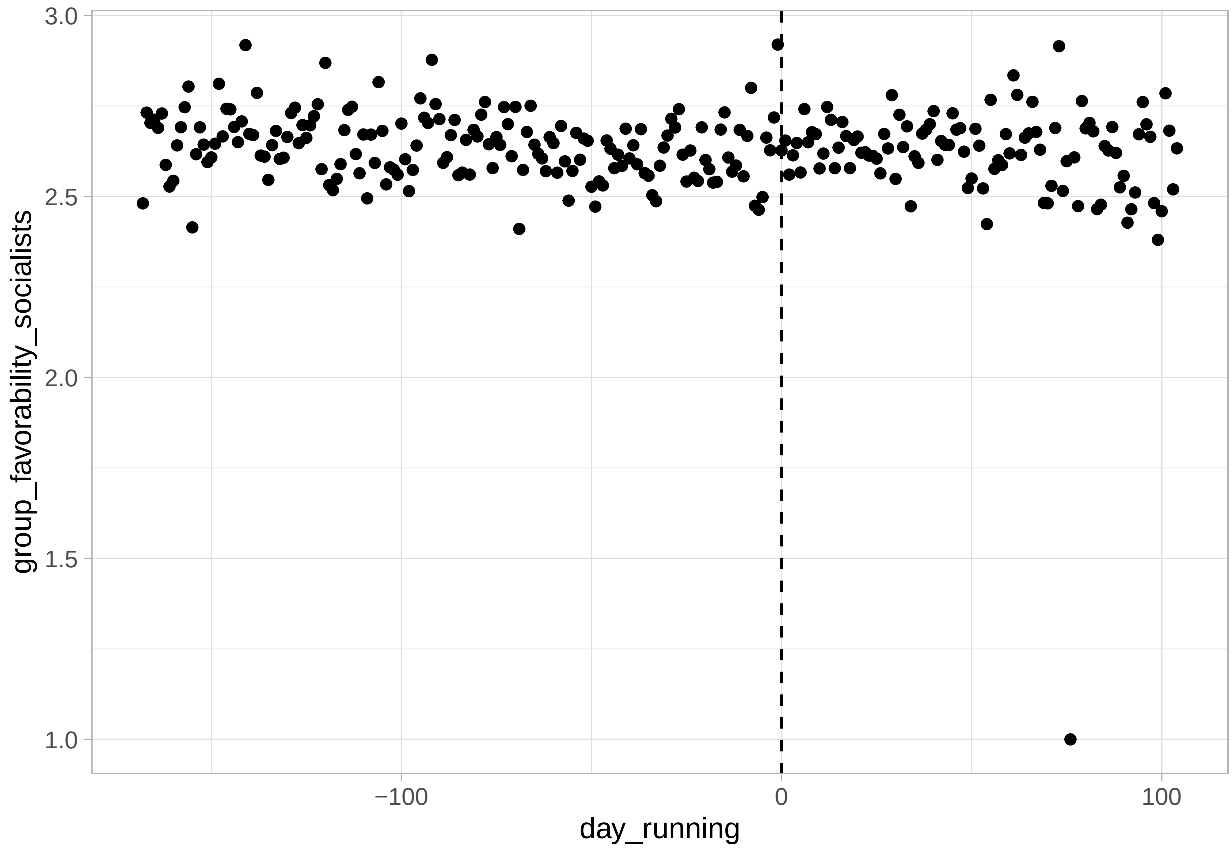
```
## [1] 2.643791
```

```r
t.test(protest_df_bydate$group_favorability_socialists[protest_df_bydate$treatment > 0], protest_df_byda
```

```
##
##  Welch Two Sample t-test
##
## data:  protest_df_bydate$group_favorability_socialists[protest_df_bydate$treatment > 0] and protest_d
## t = -1.4883, df = 104432, p-value = 0.1367
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.021416061  0.002929492
## sample estimates:
## mean of x mean of y
##  2.634548  2.643791
```

```r
# With a p-value of 0.1367, we can't reject the null hypothesis that the means
# of group_favorability_socialists before and after the treatment are equal.
```

```r
# Visualizing group_faovrability_socialists

protest_df_bydate %>%
  group_by(day, treatment, day_running) %>%
  summarize(group_favorability_socialists = mean(group_favorability_socialists), .groups = "drop") %>%
  ggplot(aes(x = day_running, y = group_favorability_socialists)) +
    geom_point() +
    geom_vline(xintercept = 0, lty = "dashed") +
    theme_light()
```



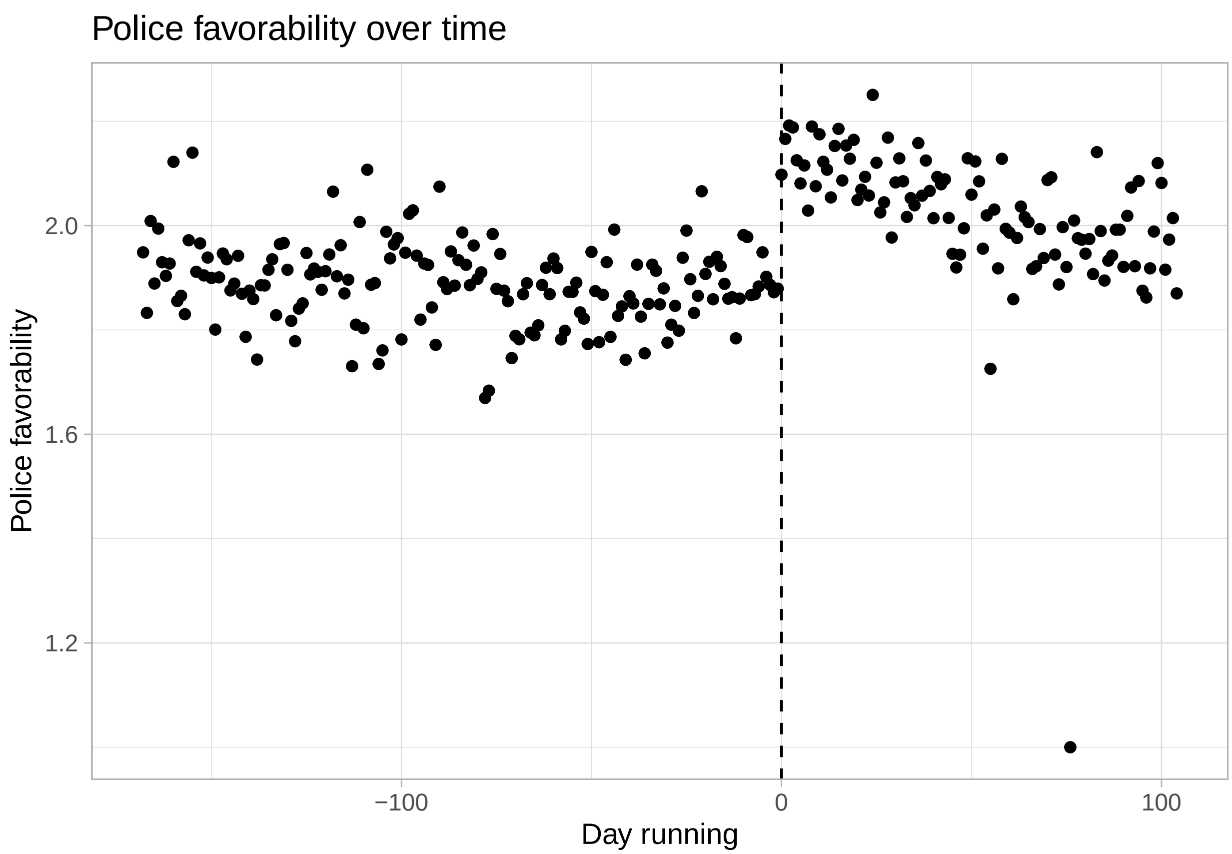
Question 4

###Part a In order to create figures similar to the panels in Figure 2 in Reny and Newman (2021) we must first manipulate the data to be more usable. If we intend to graph the average of each outcome variable for each day, on what variable should we group the data using group_by? **Create a new object that is the data split out by the appropriate group and producing the average for each of the two outcome variables for each day. Also be sure to preserve an indicator for whether the observations are from before or after the spread of the protests.**

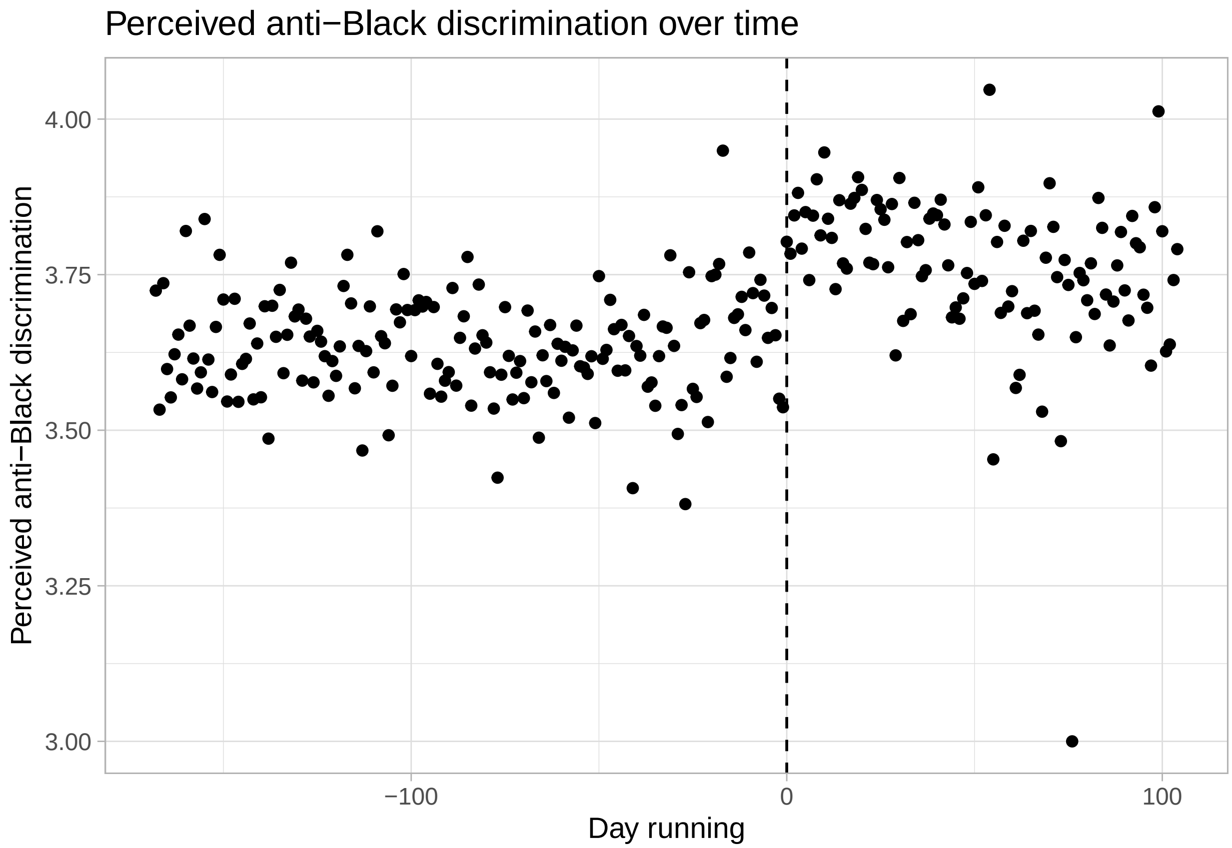
```
protest_clean_all<- protest_df_bydate %>%
  group_by(day, treatment, day_running) %>%
  summarize(police = mean(group_favorability_the_police),
            discrimination_mean = mean(discrimination_blacks), .groups = "drop")
```

###Part b Graph the results for the entire sample. **Graph the results for the entire sample for both outcome variables by day. Include a vertical line demarcating when the protests started to spread. Does there appear to be a shift in the outcome variables from before to after the protests began to spread?**

```
protest_clean_all %>%
  ggplot(aes(x = day_running, police)) +
    geom_point() +
    geom_vline(xintercept = 0, lty = "dashed") +
    theme_light() +
    labs(x = "Day running", y = "Police favorability", title = "Police favorability over time")
```



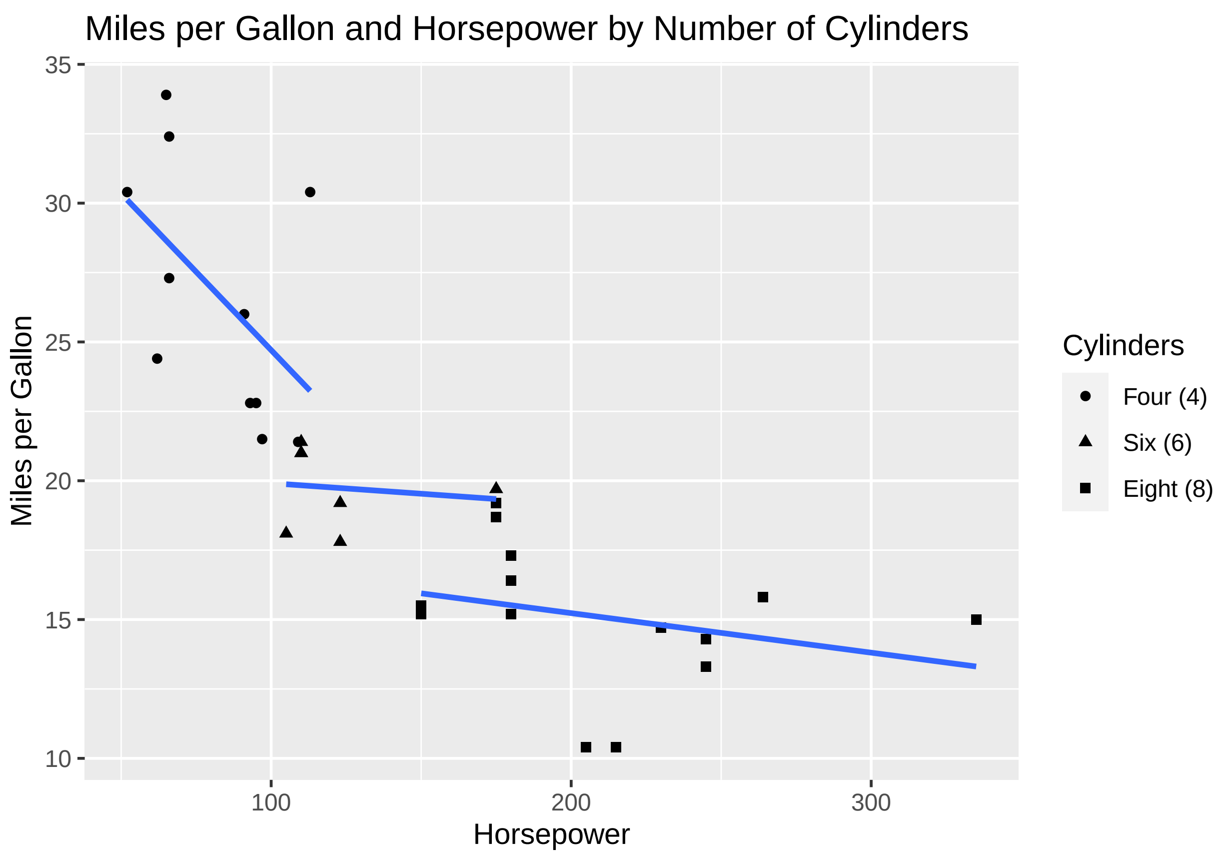

```
protest_clean_all %>%
  ggplot(aes(x = day_running, discrimination_mean)) +
    geom_point() +
    geom_vline(xintercept = 0, lty = "dashed") +
    theme_light() +
    labs(x = "Day running", y = "Perceived anti-Black discrimination", title = "Perceived anti-Black di
```



###Part c It might be useful to more clearly illustrate the differences in the trend lines before and after the protests began. **Modify the code below to include a separate line of best fit for before and after the protests began. Does the trend line align with your previous reading of the graph? Remember to add a vertical line demarcating for the onset of treatment.**

```
#An example of how to do multiple lies of best fit using example data from mtcars (mtcars is a dataset

ggplot(data=mtcars, aes(x=hp, y = mpg, shape=as.factor(cyl))) +
  geom_point() +
  geom_smooth(method="lm", se=FALSE) +
  scale_shape_discrete("Cylinders", labels=c("Four (4)", "Six (6)", "Eight (8)")) +
  ggtitle("Miles per Gallon and Horsepower by Number of Cylinders") +
  xlab("Horsepower") +
  ylab("Miles per Gallon")
```

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

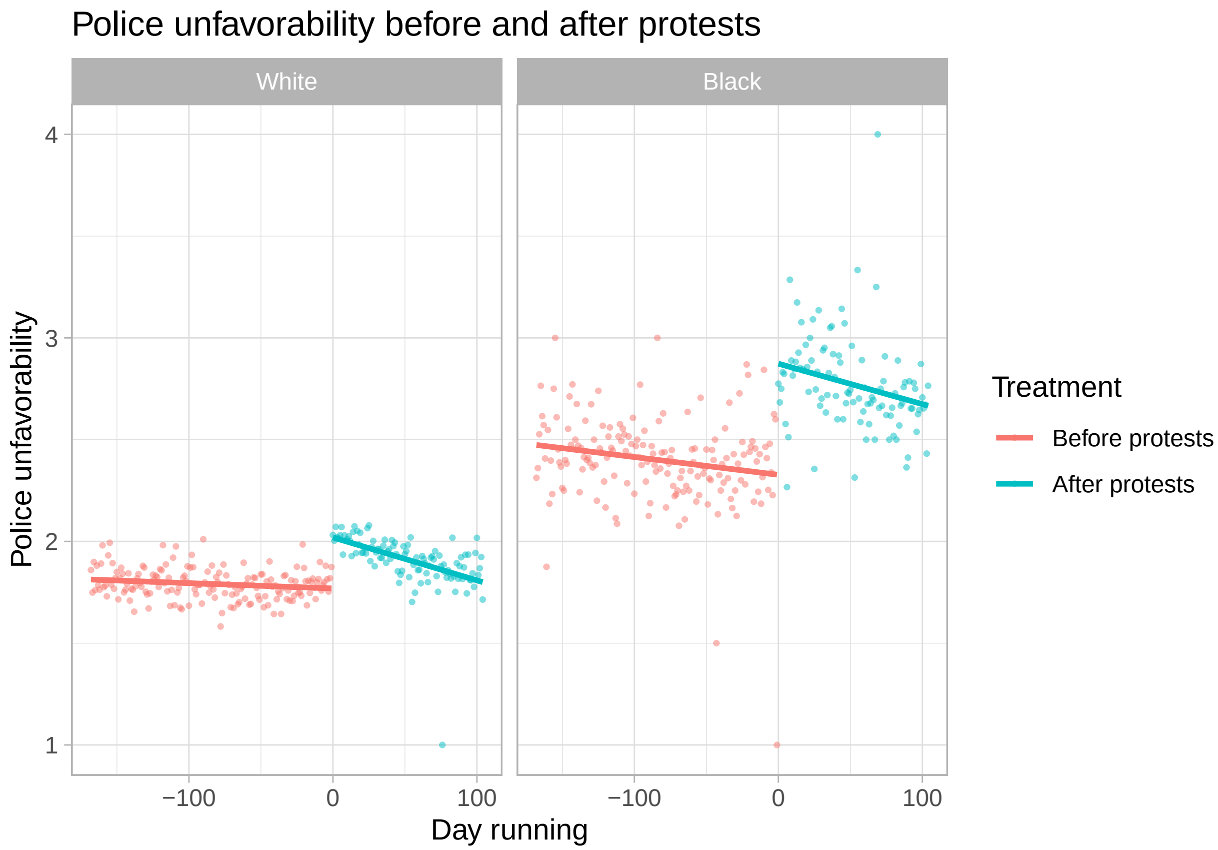


Question 5

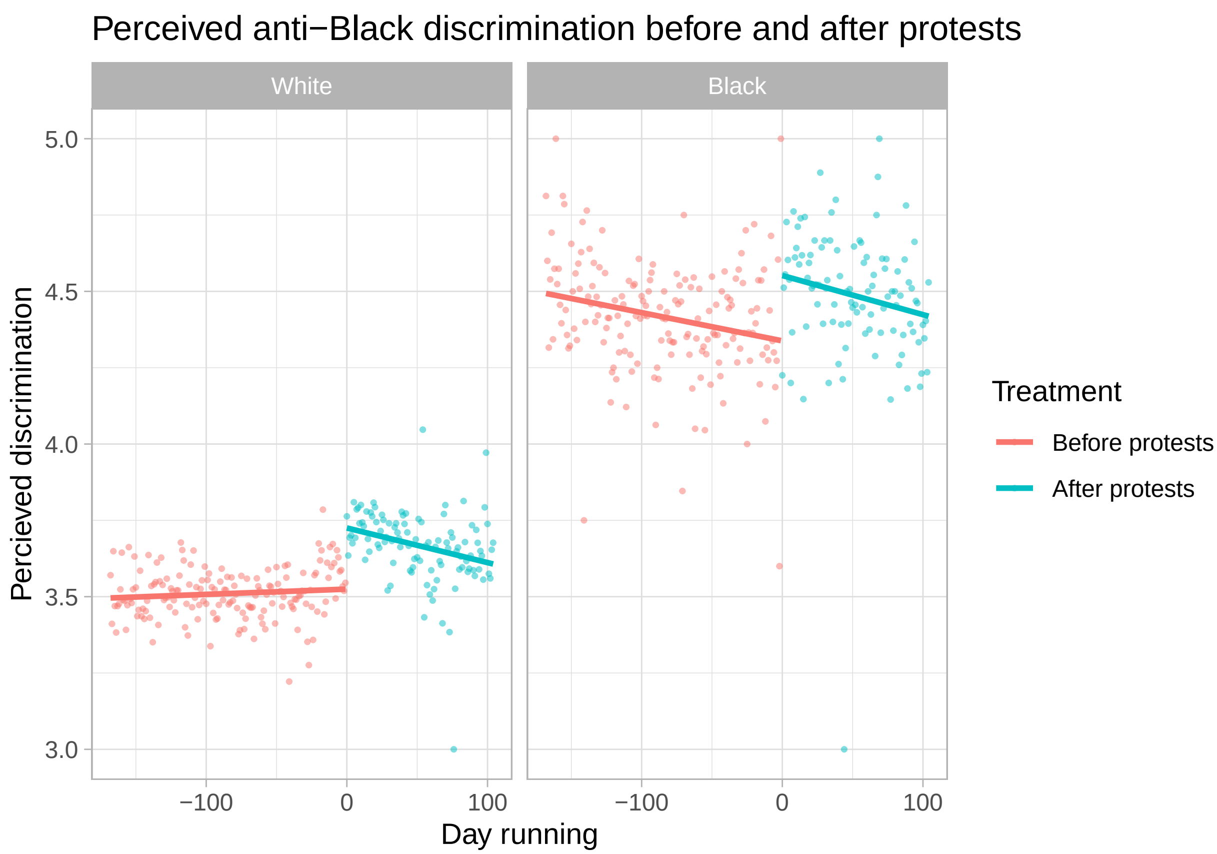
###Part a The attitudes in question are no doubt highly influenced by the respondent's race and ethnicity. How do the graphs from question 4 differ for white and Black respondents. **Subset the data to include only white respondents and recreate the graphs from part c of question 4. Do the same with the data from only Black respondents. How do these differ from each other? Hint: Be careful when subsetting white responses to not also include Hispanic responses.**

```
protest_clean_race <- protest_df_bydate %>%
  group_by(day, treatment, day_running, race_ethnicity) %>%
  summarize(police = mean(group_favorability_the_police),
            discrimination_mean = mean(discrimination_blacks), .groups = "drop")

labels <- c("1" = "White", "2" = "Black")

protest_clean_race %>%
  filter(race_ethnicity == 1 | race_ethnicity == 2) %>%
  ggplot(aes(x = day_running, y = police, color = as.factor(treatment))) +
    geom_point(size = 0.5, alpha = 0.5) +
    geom_smooth(method = "lm", formula = y ~ x, se = F) +
    facet_wrap("race_ethnicity", labeller = as_labeller(labels)) +
    theme_light() +
    scale_color_discrete(name = "Treatment", labels = c("0" = "Before protests", "1" = "After protests"
    labs(title = "Police unfavorability before and after protests",
```

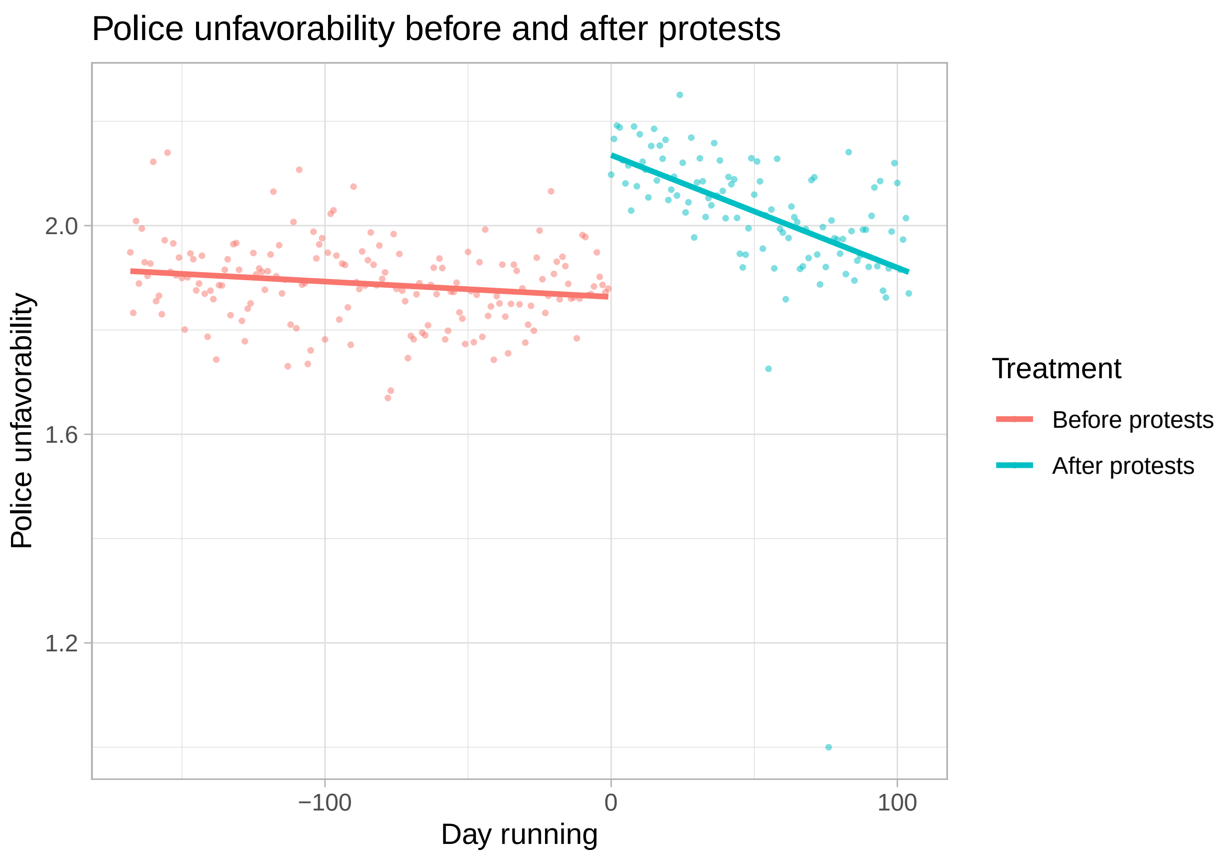
```
       x = "Day running",
       y = "Police unfavorability")
```




```
protest_clean_race %>%
  filter(race_ethnicity == 1 | race_ethnicity == 2) %>%
  ggplot(aes(x = day_running, y = discrimination_mean, color = as.factor(treatment))) +
    geom_point(size = 0.5, alpha = 0.5) +
    geom_smooth(method = "lm", formula = y ~ x, se = F) +
    facet_wrap("race_ethnicity", labeller = as_labeller(labels)) +
    theme_light() +
    scale_color_discrete(name = "Treatment", labels = c("0" = "Before protests", "1" = "After protests"
    labs(title = "Perceived anti-Black discrimination before and after protests",
         x = "Day running",
         y = "Percieved discrimination")
```

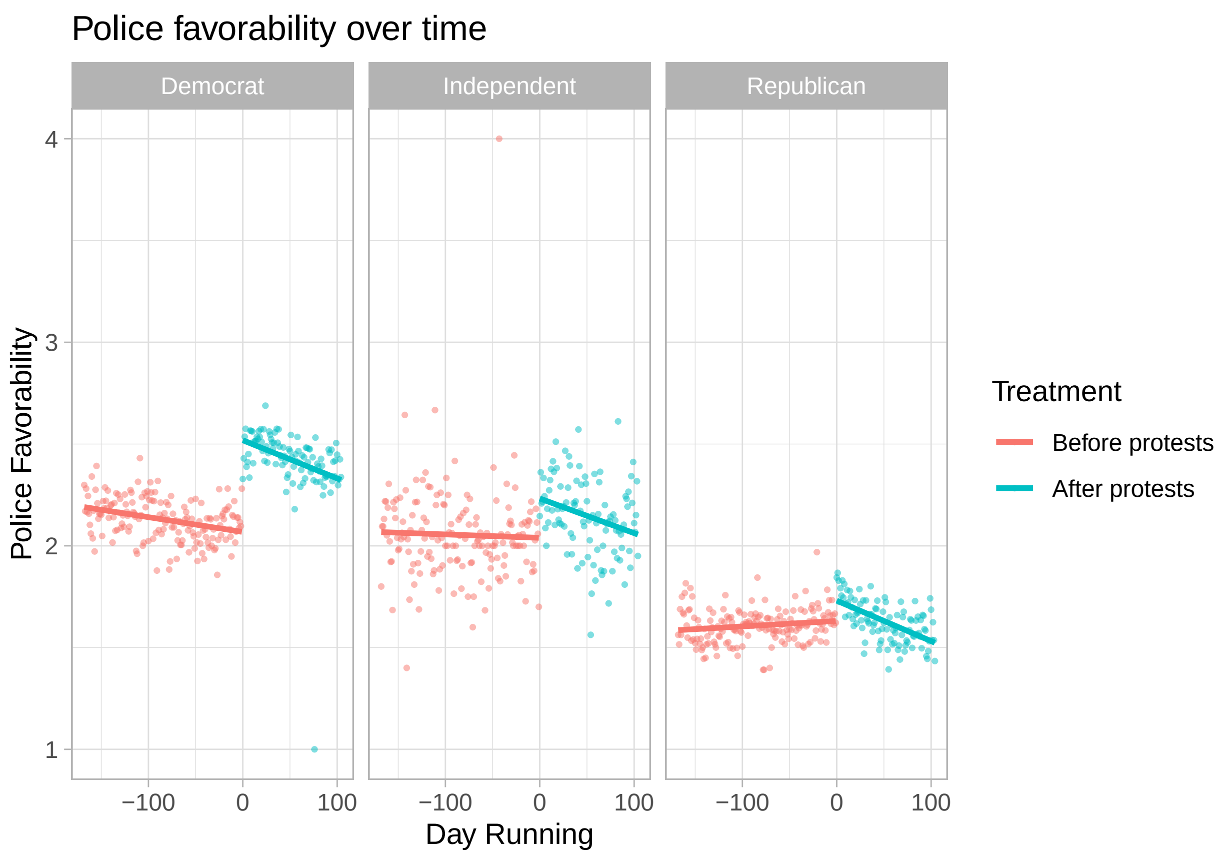



```
protest_clean_all %>%
  ggplot(aes(x = day_running, y = police, color = as.factor(treatment))) +
    geom_point(size = 0.5, alpha = 0.5) +
    geom_smooth(method = "lm", formula = y ~ x, se = F) +
    theme_light() +
    scale_color_discrete(name = "Treatment", labels = c("0" = "Before protests", "1" = "After protests"
    labs(title = "Police unfavorability before and after protests",
         x = "Day running",
         y = "Police unfavorability")
```

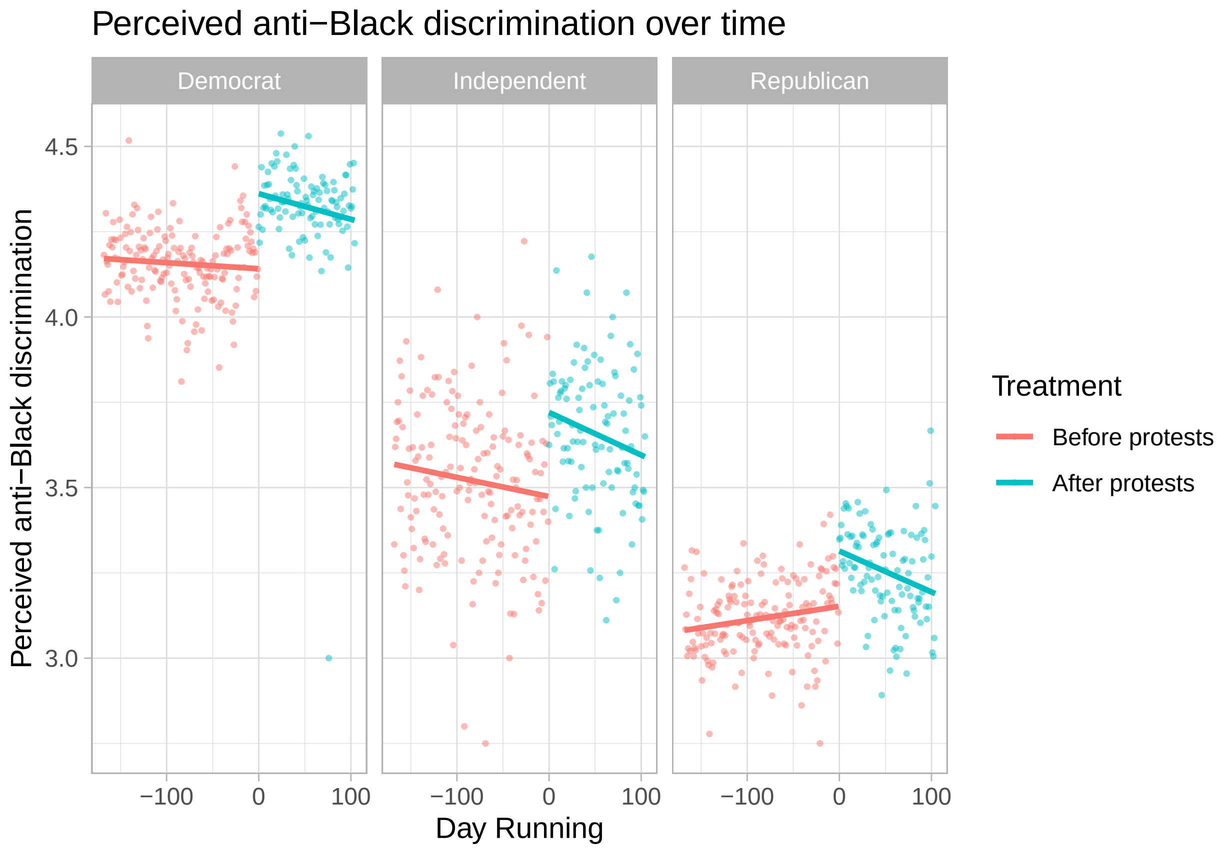


###Part b As we have learned partisanship heavily influences how people take in and process new information. **Split the sample into Democrats, Republicans and independents and use them to produce the same graphs as part a (either all in the same figure or separate). Compare both the level and the trends for each party affiliation. What could this imply about how partisanship affects processing?**

```
protest_df_bydate %>%
  group_by(pid7) %>%
  mutate(party = case_when(pid7 < 4 ~ "Democrat",
                           pid7 == 4 ~ "Independent",
                           T ~ "Republican")) %>%
  ungroup() %>%
  group_by(day, treatment, day_running, party) %>%
  summarize(police = mean(group_favorability_the_police),
            discrimination_mean = mean(discrimination_blacks), .groups = "drop") %>%
  ggplot(aes(x = day_running, y = police, color = as.factor(treatment))) +
    geom_point(size = 0.5, alpha = 0.5) +
    geom_smooth(method = "lm", formula = y ~ x, se = F) +
    facet_wrap("party") +
    theme_light() +
    scale_color_discrete(name = "Treatment", labels = c("0" = "Before protests", "1" = "After protests"
    labs(x = "Day Running", y = "Police Favorability", title = "Police favorability over time")
```




```
protest_df_bydate %>%
  group_by(pid7) %>%
  mutate(party = case_when(pid7 < 4 ~ "Democrat",
                           pid7 == 4 ~ "Independent",
                           T ~ "Republican")) %>%
  ungroup() %>%
  group_by(day, treatment, day_running, party) %>%
  summarize(police = mean(group_favorability_the_police),
            discrimination_mean = mean(discrimination_blacks), .groups = "drop") %>%
  ggplot(aes(x = day_running, y = discrimination_mean, color = as.factor(treatment))) +
    geom_point(size = 0.5, alpha = 0.5) +
    geom_smooth(method = "lm", formula = y ~ x, se = F) +
    facet_wrap("party") +
    theme_light() +
    scale_color_discrete(name = "Treatment", labels = c("0" = "Before protests", "1" = "After protests"
    labs(x = "Day Running", y = "Perceived anti-Black discrimination",
         title = "Perceived anti-Black discrimination over time")
```



##Question 6:

###Part a The graphs in questions 4 and 5 indicate that the effects dissipate as time progresses past the onset of the protests. **Explain why that might be the case? What does this indicate about whether or not attitudes towards the police are symbolic or not?**

###Part b One way to look at the effect decay is to bin the post-protest data and compare averages. **Split the post-protest data into however many groups you choose and compare the period directly after the protest with the latest period in the data. What are the differences in means for the outcomes?**

Question 7

###Part a What are some reasons we might be unconvinced by the comparison of aggregate survey results from a time before and after an event? Do you think they apply here?

###Part b There is often a problem in conducting surveys of non-response bias. That is, the people who answer surveys may differ from the people who do not answer surveys and the differences may vary over time. This is especially damaging to inference when non-response is correlated with the outcomes being measured. 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. Test whether this is the case in the Reny and Newman data. **Test whether there is balance between the respondents before and after the onset of the protests along two demographic traits that you would expect to correlate with the measured responses to the outcome variables.**

###Part c Racial resentment is often considered a symbolic attitude in strength and consistency. Examine the before and after levels of racial resentment as measured by the question from the racial resentment scale about the impact of generations of slavery and discrimination (racial_attitudes_generations). **Graph the average racial_attitudes_generations (remember the direction of how it is coded!) by day like other outcome variables. Does it behave like the other outcome variables? Does the data support that racial attitudes are symbolic attitudes?**

Question 8: Data Science Question

###Part a Run an initial regression examining the relationship between favorability towards the police, party, and treatment. **Run a regression examining party and the onset of the protests' effect on favorability towards the police. Interpret the results**

```
m1 <- lm(group_favorability_the_police ~ pid7 + treatment, data = protest_df_bydate)
summary(m1)
```

```
##
## Call:
## lm(formula = group_favorability_the_police ~ pid7 + treatment,
##     data = protest_df_bydate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4203 -0.6591 -0.2625  0.5797  2.4986
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.389334   0.005504  434.14   <2e-16 ***
## pid7        -0.126851   0.001098 -115.53   <2e-16 ***
## treatment    0.157766   0.005388   29.28   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9361 on 126772 degrees of freedom
## Multiple R-squared:  0.0999, Adjusted R-squared:  0.09989
## F-statistic:  7035 on 2 and 126772 DF,  p-value: < 2.2e-16
```

The regression indicates that party and treatment are both significantly impact individuals' favorability towards the police. The coefficient for pid7 is -0.13, meaning as respondents become more Republican (higher pid7 value), their favorability towards the police increases (lower group_favorability_the_police score). On the flip side, when respondents are more Democrat (lower pid7 value), their favorability towards the police decreases (higher group_favorability_the_police score). Additionally, the effect of treatment is significant – after the protests (treatment = 1) favorability towards the police decreases.

###Part b The above functional form probably does not accurately model the relationship of all the relevant covariates in the dataset. What functional form would you recommend using and why? What covariates would you add? Is there need for an interaction term? **Run a regression of your specificiation and interpret the results. Justify your choices in modeling.**

```
m2 <- lm(group_favorability_the_police ~ ideo5 + treatment, data = protest_df_bydate)
m3 <- lm(group_favorability_the_police ~ ideo5 + pid7 + treatment, data = protest_df_bydate)

stargazer(m1, m2, m3, type = "text")
```

```
##
## ==========================================================================================================
##                                                          Dependent variable:
##                     ------------------------------------------------------------------------------------
##                                                   group_favorability_the_police
##                               (1)                             (2)                             (3)
## ----------------------------------------------------------------------------------------------------------
## pid7                      -0.127***                                                       -0.083***
##                            (0.001)                                                         (0.001)
##
## ideo5                                                     -0.248***                       -0.153***
##                                                            (0.002)                         (0.003)
##
## treatment                  0.158***                        0.145***                        0.153***
##                            (0.005)                         (0.005)                         (0.005)
##
## Constant                   2.389***                        2.634***                        2.677***
##                            (0.006)                         (0.007)                         (0.007)
##
## ----------------------------------------------------------------------------------------------------------
## Observations               126,775                         126,775                         126,775
## R2                          0.100                           0.096                           0.123
## Adjusted R2                 0.100                           0.096                           0.123
## Residual Std. Error      0.936 (df = 126772)             0.938 (df = 126772)             0.924 (df = 126
## F Statistic         7,035.005*** (df = 2; 126772) 6,696.951*** (df = 2; 126772) 5,929.703*** (df = 3
## ==========================================================================================================
## Note:                                                                               *p<0.1; **p<0.05; 
```

First I would switch party identification for political ideology, since I think ideology is likely what determines party identification, and ideology probably has a stronger impact on how you view the world. Analyzing the data using ideology instead cuts out the ambiguity of which factors are determining party identification. Furthermore, comparing regression models using pid7 or ideo5 shows that ideo5 has a stronger effect on police favorability, which I think makes it a more informative variable to include in the model.

I would also use an interaction term ideo5*treatment, since I think it's likely your ideology impacts how you perceived the George Floyd protests. The charts in question 5b show that the 1. Democrats have lower pre-protest favorability towards police than Republicans, and 2. Democrats exhibit a greater post-protest decrease in favorability than Republicans.

```r
m4 <- lm(group_favorability_the_police ~ ideo5 + treatment + ideo5*treatment, data = protest_df_bydate)
summary(m4)
```

```
##
## Call:
## lm(formula = group_favorability_the_police ~ ideo5 + treatment +
##     ideo5 * treatment, data = protest_df_bydate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6343 -0.7361 -0.1026  0.5633  2.5633
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.529204   0.009227  274.11   <2e-16 ***
```

```
## ideo5           -0.213285   0.002851  -74.82   <2e-16 ***
## treatment        0.404554   0.014547   27.81   <2e-16 ***
## ideo5:treatment -0.086132   0.004484  -19.21   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.937 on 126771 degrees of freedom
## Multiple R-squared:  0.09818,    Adjusted R-squared:  0.09816
## F-statistic:  4601 on 3 and 126771 DF,  p-value: < 2.2e-16
```

The model shows that the interaction term is in fact significant; when ideo5 is higher (survey respondents are more conservative), the effect of treatment is lower (favorability towards police declines less).

###Part c Linear models are not well suited for bounded ordinal responses. Instead ordinal logit or probit models are frequently employed in order to capture a) that the outcomes are restricted to a scale (in the case of police unfavorability 1-4) and b) that the differences between different rungs on the scale are not necessarily equivalent (going from very unfavorable to somewhat unfavorable is not necessarily the same difference as going from somewhat unfavorable to somewhat favorable). **Using the code below from the MASS package run an ordinal probit model using the same model as part b. How do the coefficients differ from part b?**

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
select <- dplyr::select

protest_probit_data <- protest_df_bydate %>%
  mutate(group_favorability_the_police = as_factor(group_favorability_the_police))

m5c <- polr(formula = group_favorability_the_police ~ ideo5 + treatment + ideo5*treatment,
     data = protest_probit_data,
     method = "probit")

summary(m5c)
```

```
##
## Re-fitting to get Hessian
```

```
## Call:
## polr(formula = group_favorability_the_police ~ ideo5 + treatment +
##     ideo5 * treatment, data = protest_probit_data, method = "probit")
##
## Coefficients:
##                     Value Std. Error t value
## ideo5           -0.26468   0.003474  -76.19
## treatment        0.41993   0.017175   24.45
```

```
## ideo5:treatment -0.09031   0.005450  -16.57
##
## Intercepts:
##                                             Value    Std. Error t value
## Very favorable|Somewhat favorable            -0.9706   0.0112   -86.2815
## Somewhat favorable|Somewhat unfavorable      -0.0338   0.0110    -3.0610
## Somewhat unfavorable|Very unfavorable         0.6121   0.0113    54.1692
## Very unfavorable|Not asked in this wave       6.5401   0.0113   578.7902
## Not asked in this wave|Haven't heard enough   7.0791   0.0113   626.4902
##
## Residual Deviance: 303846.86
## AIC: 303862.86
```

```r
stargazer(m5c, m4, type = "text")
```

```
##
## ===============================================================
##                               Dependent variable:
##                     -------------------------------------------
##                          group_favorability_the_police
##                      ordered                OLS
##                      probit
##                        (1)                  (2)
## ---------------------------------------------------------------
## ideo5               -0.265***            -0.213***
##                      (0.003)              (0.003)
##
## treatment           0.420***             0.405***
##                      (0.017)              (0.015)
##
## ideo5:treatment     -0.090***            -0.086***
##                      (0.005)              (0.004)
##
## Constant                                 2.529***
##                                           (0.009)
##
## ---------------------------------------------------------------
## Observations         126,775              126,775
## R2                                         0.098
## Adjusted R2                                0.098
## Residual Std. Error               0.937 (df = 126771)
## F Statistic                  4,600.561*** (df = 3; 126771)
## ===============================================================
## Note:                          *p<0.1; **p<0.05; ***p<0.01
```

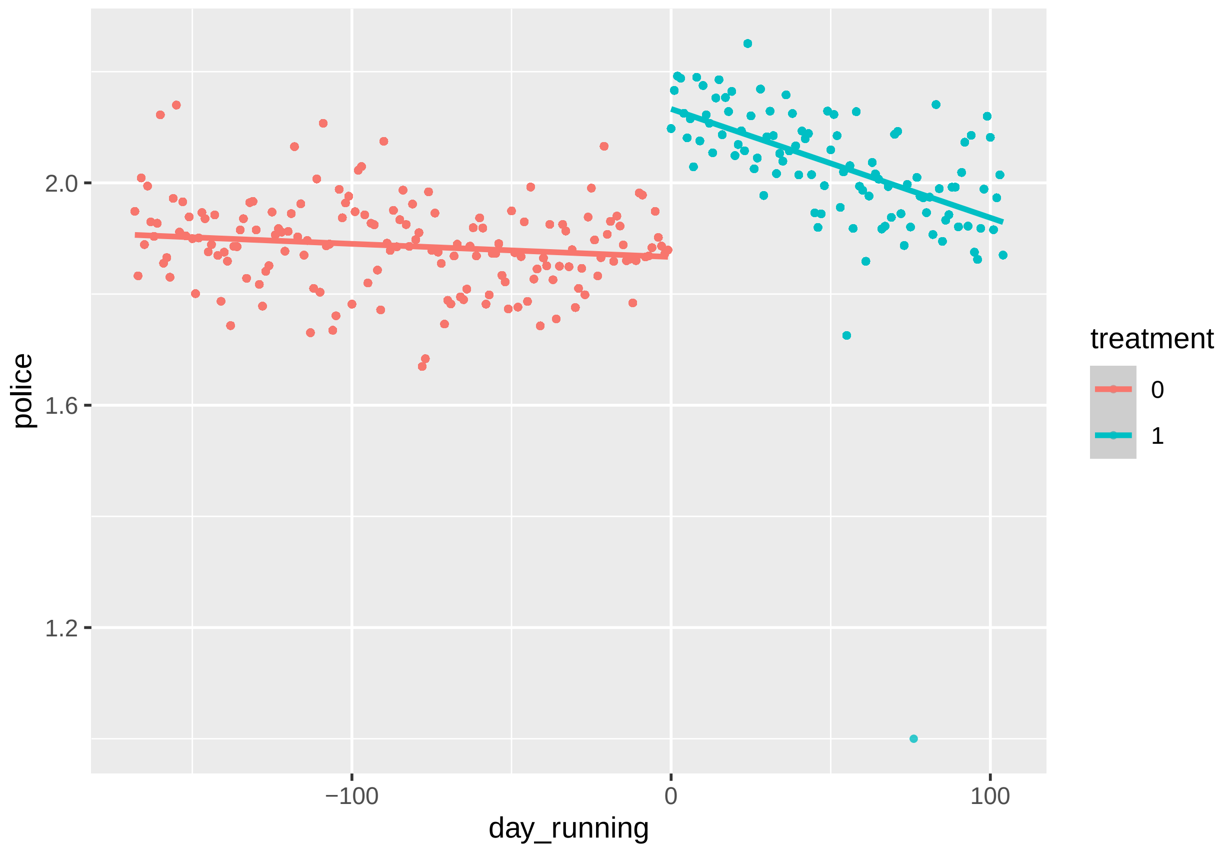
The coefficients for the linear and pobit models are relative similar, but the coefficients for the probit model are all slightly higher (around 0.01-0.05 higher). The probit model also lists several intercept values (which I don't know how to interpret).

###Section

```r
protest_df_bydate %>%
  group_by(day_running) %>%
```

```r
  drop_na() %>%
  summarize(police = mean(group_favorability_the_police),
            disc = mean(discrimination_blacks),
            treatment = as.factor(treatment))%>%
  ggplot(aes(x = day_running, y = police, color = treatment)) +
    geom_point(size = 0.8, alpha = 0.8) +
    geom_smooth(method = "lm", formula = y ~ x)
```

```
## `summarise()` regrouping output by 'day_running' (override with `.groups` argument)
```



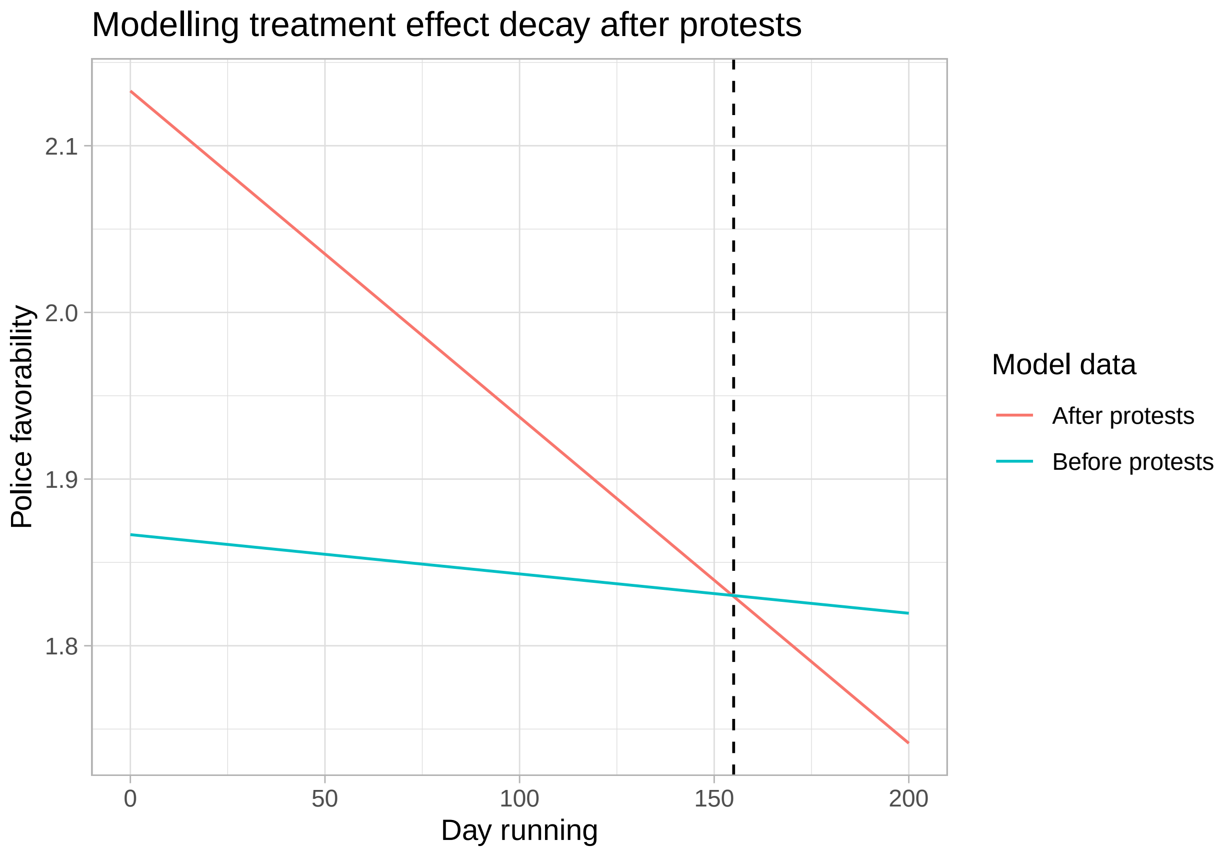

```r
decay_before <- protest_df_bydate %>%
  group_by(day_running) %>%
  drop_na() %>%
  summarize(police = mean(group_favorability_the_police),
            disc = mean(discrimination_blacks),
            treatment = as.factor(treatment), .groups = "drop") %>%
  filter(day_running < 0)

decay_after <- protest_df_bydate %>%
  group_by(day_running) %>%
  drop_na() %>%
  summarize(police = mean(group_favorability_the_police),
            disc = mean(discrimination_blacks),
            treatment = as.factor(treatment), .groups = "drop") %>%
```

```r
  filter(day_running >= 0)

m_before <- lm(police ~ day_running, data = decay_before)
m_after <- lm(police ~ day_running, data = decay_after)

stargazer(m_before, m_after, type = "text")
```

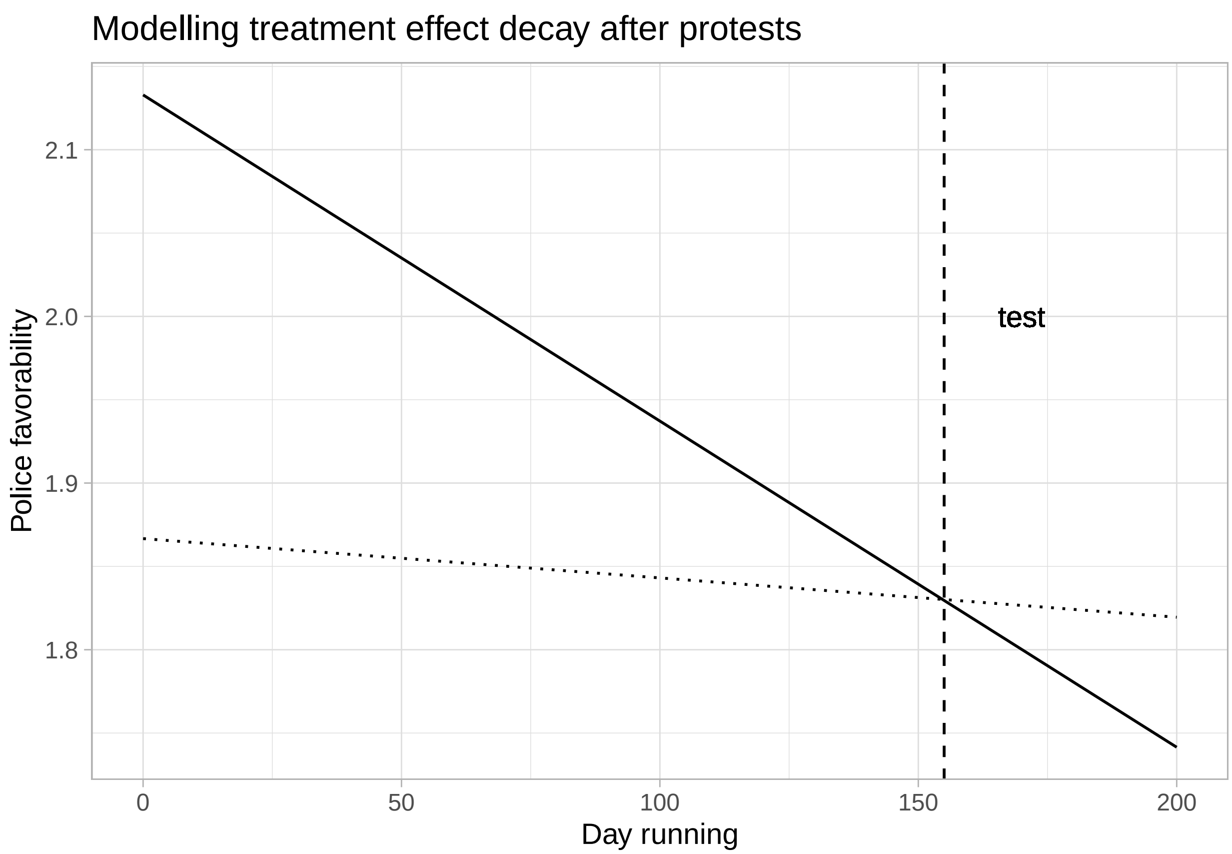
```
##
## ==============================================================================
##                                          Dependent variable:
##                      ---------------------------------------------------------
##                                                 police
##                                 (1)                            (2)
## ------------------------------------------------------------------------------
## day_running                 -0.0002***                     -0.002***
##                             (0.00001)                      (0.00001)
##
## Constant                     1.867***                       2.133***
##                             (0.0005)                        (0.001)
##
## ------------------------------------------------------------------------------
## Observations                  77,157                         49,618
## R2                            0.027                          0.418
## Adjusted R2                   0.027                          0.418
## Residual Std. Error     0.069 (df = 77155)             0.070 (df = 49616)
## F Statistic         2,162.321*** (df = 1; 77155) 35,572.290*** (df = 1; 49616)
## ==============================================================================
## Note:                                              *p<0.1; **p<0.05; ***p<0.01
```

```r
pred_data <- data.frame(day_running = seq(0, 200, 1))

preds_before <- predict(m_before, pred_data, se.fit = T)
preds_before <- tibble(police = preds_before$fit,
                       day_running = seq(0, 200, 1),
                       treatment = 0)

preds_after <- predict(m_after, pred_data, se.fit = T)
preds_after <- tibble(police = preds_after$fit,
                      day_running = seq(0, 200, 1),
                      treatment = 1)

preds_combined <- rbind(preds_before, preds_after) %>%
  mutate(treatment = ifelse(treatment == 0, "before", "after")) %>%
  pivot_wider(names_from = treatment, values_from = police) %>%
  mutate(diff = before - after)

preds_combined %>%
  # mutate(intercept = 155) %>%
  select(-diff) %>%
  pivot_longer(cols = before:after, values_to = "value", names_to = "type") %>%
  ggplot(aes(x = day_running, y = value, color = type)) +
    geom_path() +
    geom_vline(xintercept = 155, lty = "dashed") +
```

```
    labs(x = "Day running", y = "Police favorability",
         title = "Modelling treatment effect decay after protests") +
    scale_color_discrete(name = "Model data",
                         labels = c("after" = "After protests",
                                    "before" = "Before protests")) +
    theme_light()
```




```
preds_combined %>%
  ggplot(aes(x = day_running)) +
    geom_path(aes(y = before), lty = "dotted") +
    geom_path(aes(y = after)) +
    geom_vline(xintercept = 155, lty = "dashed") +
    geom_text(aes(x = 170, y = 2), label = "test") +
    labs(x = "Day running", y = "Police favorability",
         title = "Modelling treatment effect decay after protests") +
    theme_light()
```


Modelling treatment effect decay after protests
Police favorability
2.1
2.0
1.9
1.8
test
0
50
100
150
200
Day running