

Data Exploration: Contextual Influences

Patrick Adams

November 11, 2021

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

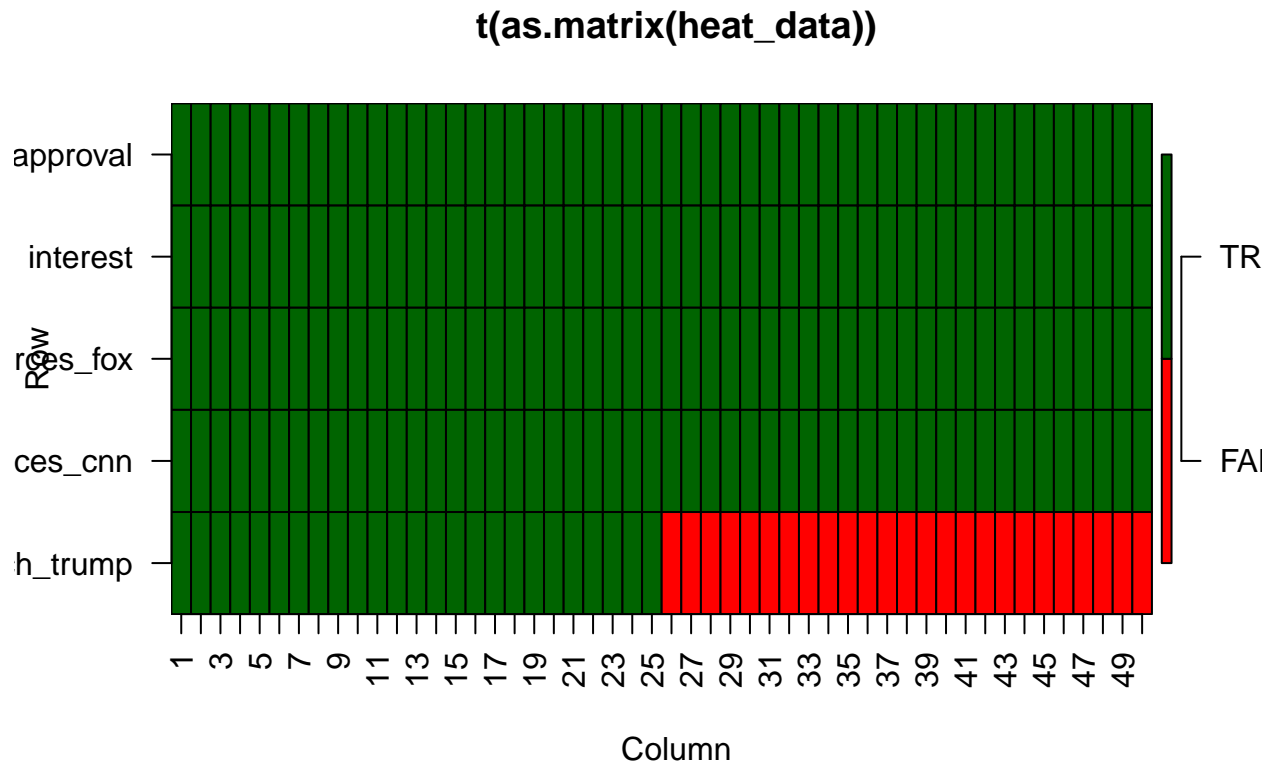
```
NationscapeVars <- full_join(NationscapeVars_1,NationscapeVars_2) %>% # the full_join function keeps all  
  replace(is.na(.),0) # since we know that the NAs generated in the last command weren't asked in the w
```

This week we again were exploring the Nationscape dataset (Tausanovitch and Vavreck 2020), and the question I wanted to answer with this weeks data blog was:

How did different news source coverage of President Trump's 2019/20 impeachment trial affect his approval ratings among their viewers?

My hypothesis is that those who view CNN are more liberal than those who view FOX, and thus they will have lower approval ratings of Trump to begin with. However throughout the trial, depending on the coverage provided, the approval ratings of the President Trump would have fallen more amongst CNN viewers due to less favourable coverage than that which FOX news gave.

```
heat_data <- NationscapeVars %>% mutate(across(.cols = everything(), as.logical)) %>% select(c(pres_appr  
plot(t(as.matrix(heat_data)), col = c('red','darkgreen'), las = 2) # we're using the plot() function fr
```



This plot helps show the availability of the data points I chose to look at throughout the overall data set.

The feedback I received was that the biases held by the separate viewership may impact the change seen; for example, as I touched on earlier those who watch CNN are potentially predisposed to a lower approval ratings of Trump, and so any change is more easily seen. Equally there is scenario in which, as this is measured on a four point scale, CNN viewers already hold such a low approval rating that they cannot get much worse and thus the ratings given by FOX viewers will potentially fall to a more significant degree.

Some prior literature on this topic shows just how divided the viewership of different outlets can be:

Martin, G. J., & Yurukoglu, A. (2017). Bias in Cable News. The American Economic Review, 107(9), 2565–2599. <https://doi.org/10.1257/aer.20160812>

Hyun, K. D., & Moon, S. J. (2016). Agenda Setting in the Partisan TV News Context. Journalism & Mass Communication Quarterly, 93(3), 509–529. <https://doi.org/10.1177/1077699016628820>

```
file_names_1 <- list.files("Nationscape-DataRelease_WeeklyMaterials_DTA/phase_1_v20200814/") %>% .[1:24]

phase_1 <- map_dfr(.x = file_names_1,
  ~read_dta(file = str_c("Nationscape-DataRelease_WeeklyMaterials_DTA/phase_1_v20200814/", .x))

TI <- phase_1 %>%
  mutate(across(.cols = everything(), ~na_if(., 999))) %>%
  mutate(across(.cols = everything(), ~na_if(., 888)))

TI <- TI %>%
  mutate(treated = if_else(start_date > as.Date('2019-12-18'), TRUE, FALSE))
```

```
t.test(pres_approval ~ treated, data = TI)
```

```
##
## Welch Two Sample t-test
##
## data: pres_approval by treated
## t = 3.5674, df = 18232, p-value = 0.0003614
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.01705904 0.05866497
## sample estimates:
## mean in group FALSE mean in group TRUE
## 2.794456 2.756594
```

```
cohen.d(pres_approval ~ treated, data = TI)
```

```
## Warning in cohen.d.formula(pres_approval ~ treated, data = TI): Cohercing rhs of
## formula to factor
```

```
##
## Cohen's d
##
## d estimate: 0.03125638 (negligible)
## 95 percent confidence interval:
## lower upper
## 0.01433310 0.04817966
```

To establish if there was a statistically significant change in Trump's approval rating before and after the December 18th Impeachment I ran a t-test on the phase 1 data. Even though there is a statistically significant change based on $p\text{-value} = 0.0003614$, the cohen's d effect size is negligible at 0.03125638. So while there is an effect, the effect size is negligible. This may be because the survey size is so large, and those surveyed each week are different, that the significance may be down to sampling error instead of a shift in the population. This is just a quick test, and can be explored further.

Producing a similar graph to those we used to visualize data in week 7 doesn't give us a particularly conclusive picture either, even if there is a slight uptick in the plotted regression line. The visualization also speaks to sampling error that I assumed after the t-test results, as can be seen by the spread of the data points.

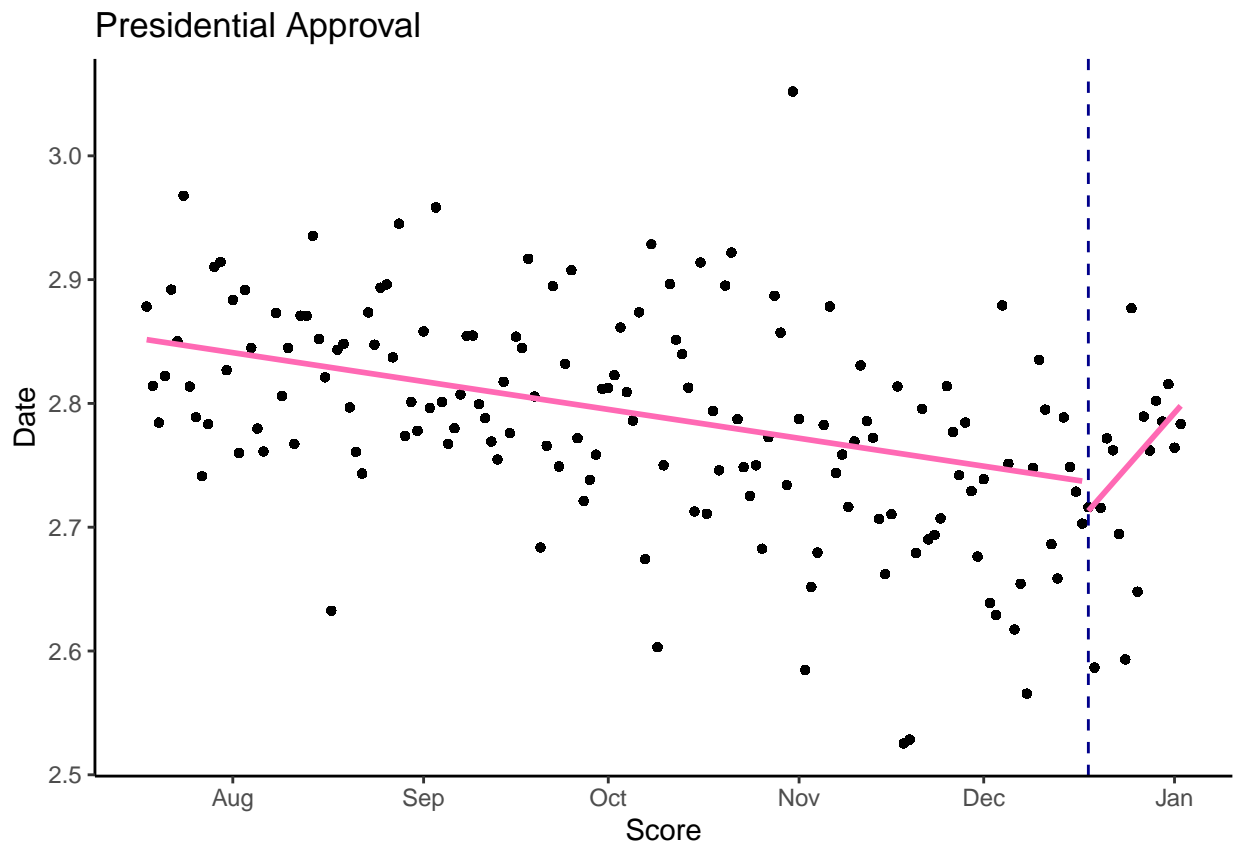
```
TI <- TI %>%
  mutate(day = as.Date(start_date, origin = "2019-12-18"))

TI_avg_by_day <- TI %>%
  group_by(day) %>%
  mutate(avg_imp = mean(impeach_trump, na.rm = T)) %>%
  mutate(avg_app = mean(pres_approval, na.rm = T))

TI_avg_by_day %>%
  group_by(day) %>%
  ggplot(aes(x = day, y = avg_app, fill = as.factor(treated))) +
  geom_point(colour = "black", size = 1) +
  geom_smooth(method="lm", se=FALSE, colour = "hotpink") +
```

```
geom_vline(xintercept = as.Date("2019-12-18"), colour = 'darkblue', linetype = "dashed") +
theme_classic() +
theme(legend.position = 'none') +
labs(title = "Presidential Approval",
      x = "Score",
      y = "Date")
```

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



```
impeach <- TI_avg_by_day %>%
  group_by(day) %>%
  ggplot(aes(x = day, y = avg_imp, fill = as.factor(treated))) +
  geom_point(colour = "black", size = 1) +
  geom_smooth(method="lm", se=FALSE, colour = "hotpink") +
  geom_vline(xintercept = as.Date("2019-12-18"), colour = 'darkblue', linetype = "dashed") +
  theme_classic() +
  theme(legend.position = 'none') +
  labs(title = "Views on Impeachment",
        x = "Score",
        y = "Date")
```

Regardless of the small likelihood of my hypotheses being correct, or at least provable from this data, I wanted to at least compare the views of the respondents who got their news from different sources. To do this I filtered out all respondents who had answered yes to viewing either CNN and no to FOX, and visa versa.

When looking at only those respondents that solely watched one news source it is easy to see how differently the viewers perceive the event. Using a linear regression model we see that those who only watch FOX are, on only a four point scale, 0.882 units more approving (as the scale goes from 1 - Strong approval through to 4 - Strong disapproval). Likewise, CNN viewers were 0.671 units less approving. A similar situation is seen with the 'Impeach Trump' data point.

Finally, to see if these partisan viewers changed their opinion I ran some more specific t-tests:

```
TI_avg_by_day %>%
  filter(fox == 1) %>%
  t.test(pres_approval ~ treated, .)

##
## Welch Two Sample t-test
##
## data: pres_approval by treated
## t = 0.81722, df = 3294, p-value = 0.4139
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.02623852 0.06374306
## sample estimates:
## mean in group FALSE mean in group TRUE
## 1.929955 1.911203
```

```
TI_avg_by_day %>%
  filter(cnn == 1) %>%
  t.test(pres_approval ~ treated, .)

##
## Welch Two Sample t-test
##
## data: pres_approval by treated
## t = 0.50384, df = 3557.5, p-value = 0.6144
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.02645884 0.04476054
## sample estimates:
## mean in group FALSE mean in group TRUE
## 3.481354 3.472203
```

As with the original t-tests, neither of these more specific tests proved to show any statistical significance. So to answer my original question, the news sources didn't change the opinions of those who only watch one or the other at all. In hindsight this makes complete sense, as based on the fact that both sets of viewers started out with such different approval ratings it shows they have probably already made up their minds and have strong biases on the subject.