Write Up

George Perrett

5/15/2020

Introduction

Since 2015 throgh today, calls to increase government regulaton of social media tech compaiins have grown. Social media companines have an inherent interest in limitting increased regulations which can disrupt many aspects of bussiness models and growth startagies.

Public opinion is an important motivator of politicians actions, and the publics veiws towards social media comaines and the extent to which they are regulated can either motivate or prevent politicans from altering existing frameworks. Political ideaology and identies are inherently related to individuls opinions about regulation and government intervention, with those identifying as more conservative favoring less government invovlement and those identifying as more liberal favoring a more active government. While political persuation is likely remains a critically important determinant in individuals views on regulating the tech secotr, individuals experiances directly interacting with social media sites and prodcuts may also influence thier views.

This analysis ustilizes Bayesain methods to test is public opinion on regulating social media compaines can be explained einterly as a function of ones political views or if the types of content and experiances users interact with on social medias sites is also associated with thier views on regulation. Put another way, are opinions towards regulating big tech all just politics, or can altering aspects of the social media expernace increase the publics favorability towards social media companines?

If opinons towards regulation are associated with aspects of the social media experiance or engaging wit specifc types of social media content, social media compaines may be able to strucute alorithmic changes that downregulate these types of content as a means of increasing thier favorability and preventing increased government intervention.

The Data

In 2018 the Pew Research Center surved a representative sample of Americans on their use or and feeling toward social media. Particiants were asked wether the government should regulate social media compaines more than they already are, about the same as they are now or less than they are now. This item was dictomized such that respondents indicatign current or lower levels of regulation were coded as being against expnading regulation and compared against those favoring expanding regulation. After dictomizing, this item was used as the outcome varible in all subseqent analyses. Favoring more regulation was coded as 1 while being againt an expansion of regulations was coded as 0.

Demographic items were respondents indicated their political identy, race and age were used as measures of political persuation and itentity. Political ideaology was descibded as either very conservative, conservative, moderate (reference class), liberal or very liberal. Age was divided by users who were 18-29, 30-49 (reference class), 50-64 or 65 and above. Respondents could indicate thier race as wither white (reference class), black, asian, mixed race or other.

User experiances with social media conrent was measured throgh three seperate variables: the frequency of encounter posts that increased negative affect, the frquency of encoutering politically charged, triggering and controversial content and the frequency of ecnountering disinformation.

Factor analysis was used to combine 3 seperate items about the frequency of seeing content that made users feel angry, depressed and lonely into a single continous latent varible representing negative affect. Simmilalry, a factor socres were used to combine a series of 4 seperate items asking about the frequency that users encountered posts about race relations, sexual harasment/assult, gun control/gun violance and immigration into a single continous latent varibale representing politically charged content. A single item where respondents were asked to descriebe if they saw more posts proting deception, saw more posts trying to point out misinformation or an equal amount each (reference class) was included as a measure of disinformation.

To understnad the relationship between political identiy, social media content and views on regulation 3 candidate models were created and compared against each other.

Hypotheses: comparing 3 seperate models

Model 1: Views on regulation as a function of political ideaology, age and race

Model 1 is the uses only political identification, race and age to predict wether a given individual factors expanding existing regulations of social media companies or beleives current regulatory practices are sufficent. If model fits the data the best, this implies that opinions on the regulation of social media compaies are an extention of already existing political beleifs.

$$regulation_i \sim Binomial(n, p_i) logit(p_i) = \alpha + \beta veryconservative_i + \beta conservative_i + \beta liberal_i + \beta veryliberal_i + \beta age: 18{-}29_i +$$

Model 2: Views on regulation as a function of political ideaology, identiy and features of social media sites

Model 2 includes the same political and identiy predictors as model 1 abut adds the measures of social media based features (negatvie affect, charged content and deceptive vs corrective information).

If model 2 fits the data better than the other models this supports the hypothesis that views towards the regulation of social media companines is not a purly political issues but is also related to the content and types of experiances encoutered by users.

$$regulation_i \sim Binomial(n, p_i) logit(p_i) = \alpha + \beta veryconservative_i + \beta conservative_i + \beta liberal_i + \beta veryliberal_i + \beta age: 18{-}29_i +$$

Model 3: Views on regulation as a function of identiy and features of social media sites with the role of user experiance varying across different political ideaologies

Model 3 allows the relationship between social media content and opinions on regulation to vary across different political ideaologies. The opinions of those with less extreame political opinions may be more mailible that those who identify as highly ideaological. By estimating different slopes for each political identiy, model 3 allows for these differential effects.If model 3 fits the data best, this implies that for some political identites, views on social media is are only a function of political identity, but for other political identies views on regulation vary with different experiances on soical media sites.

$$regulation_i \sim Binomial(n, p_i) logit(p_i) = \alpha_{ideaology_i} + \beta_{ideaology_i} negative\ affect_i + \beta_{ideaology_i} charged\ content_i + \beta_{ideaology_i} dec$$

Analysis

Selecting Priors and Drawing from the Prior Predictive Distribution

Specifying priors is the first step of a baysian analysis. Given that there already exists a large body of evidnece connecting political ideaology and views on regulation, informative priors were given to polotical identificaiton.

In contrast, less is known about relationship between age, race and social media relatred varibales. This uncertntly was reflected with less informative priors that were all centered at 0 (showing equal likelihood of favoring or not favorging increasd regulation) and given wider standard deviations.

The specific assigned priors are shown in the code block printed below. The following code not only assigns priors but draws from the prior predcitve distribution to ensure that the selected priors make plausible predictions.

```
## Set priors
# for demogrpahic only model
base_prior <- normal(c(-1, # prior for very conservative
                       -.25, # prior for conservative
                       .25, # prior for liberal
                       1, # prior for every liberal
                     rep(0, 7)),  #  weaker priors for race and age
                   c(rep(.5, 4), # tigter scale for political ideaology
                     rep(1, 7)), # more flexibility for other demogrpahics
                      autoscale = F)

## Model 1

# demographics only
M1 <- stan_glm(
  regulation ~
    very_conservative + conservative + liberal + very_liberal +
    age_18_29 + age_50_64 + age_65_up +
    black + asian + mixed_race + other,
  data = model_data,
  family = binomial,
  prior_intercept = normal(0, 2, autoscale = F),
  prior = base_prior, # informative priors for political identification
  prior_PD = T, # samples from prior predictivee distribution
  seed = 2,
  cores = 4,
)

# Model 2
## Adding priors for social media variables
full_prior <- normal(c(-1, -.25,.25, 1, # informative priors for political ideaology
                     rep(0, 11)),  #  weaker priors for race, age adn social media
                   c(rep(.5, 4), # tigter scale for political ideaology
                     rep(1, 6), # more flexibility for other demogrpahics
                     rep(.25, 2),
                     rep(1, 3)),
                      autoscale = F)

M2 <- stan_glm(
```

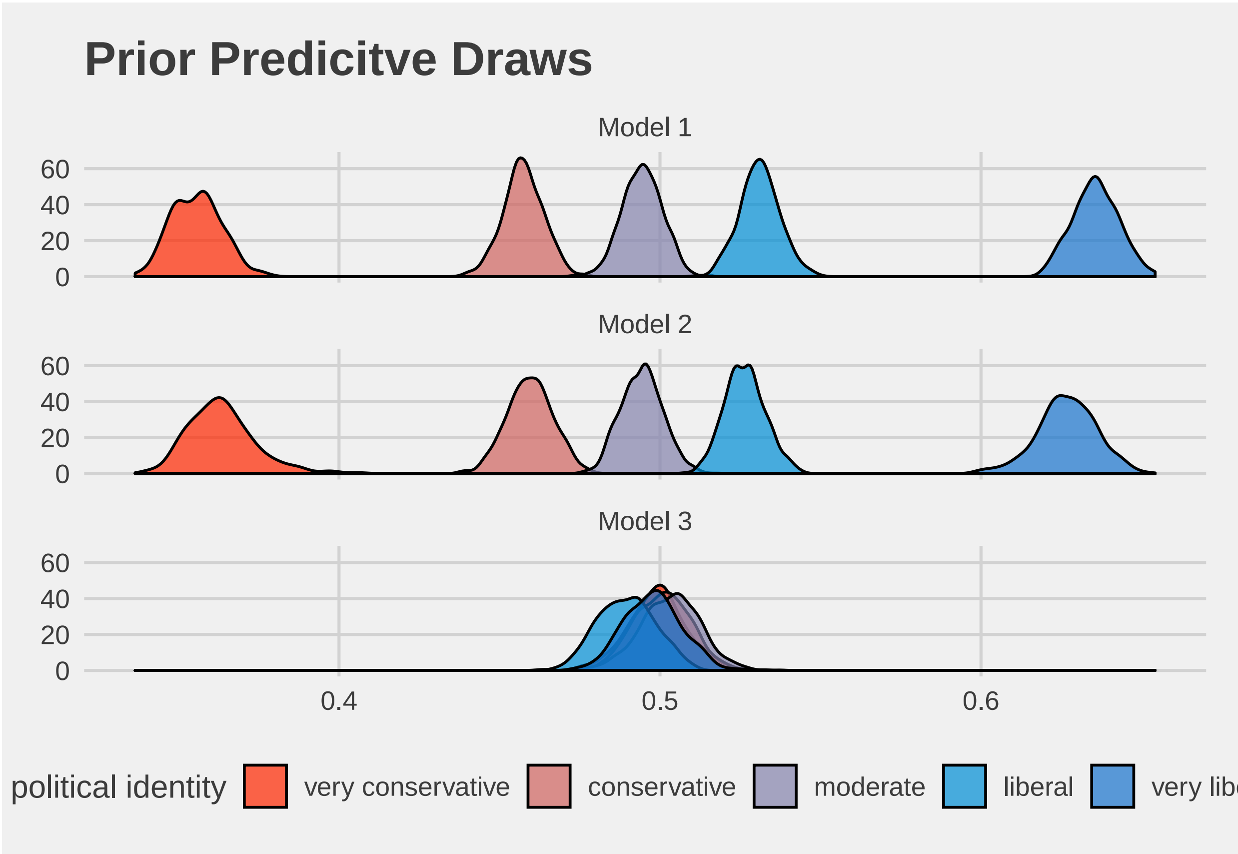
```r
  regulation ~
    very_conservative + conservative + liberal + very_liberal +
    age_18_29 + age_50_64 + age_65_up +
    black + asian + mixed_race + other +
    # adding UX varaibles
    Negative_Affect + Charged_Content + deception + correct_misinformation,
  data = model_data,
  family = binomial,
  prior_intercept = normal(0, 2, autoscale = F),
  prior = full_prior,
  prior_PD = T,
  seed = 2,
  cores = 4,
)

# Model 3
M3 <- stan_glmer(
  regulation ~
    # Population Estimates:
    Negative_Affect + Charged_Content +
    deception + correct_misinformation +
    age_18_29 + age_50_64 + age_65_up +
    black + asian + mixed_race + other +
    # Varying Slopes by Ideaology:
    (
      Negative_Affect + Charged_Content +
        deception + correct_misinformation |
        ideaology
    ),
  family = binomial("logit"),
  model_data,
  prior_intercept = normal(0, 2, autoscale = F),
  prior = normal(0,c(rep(.25, 2),rep(1.5,9)), autoscale = F),
  prior_PD = T,
  QR = TRUE,
  adapt_delta = .99,
  seed = 1234,
  chains = 2,
  cores = 2
)
```

Visulaizing the Prior Predictive Distribution

The plot shown below provides predictions drawn from the prios for each of the 3 models. Predictions are grouped by political ideaology to inspect informative prior given to differnt ideaologies.

The plot suggests that all predictions are in a reasonable range and that the assumptions of the prior specifcation are reflected in draws from the prior predcitive distribution. Different political ideaologies are grouped togehter in model 3 due to the multi-level structure of the data.



Conditioning on the Data: Fitting and Comparing Models

After assesing the prior distributions, the code chunck below conditions on the data and fits the models. These results can be used to conduct inference and evaluate the established hypotheses.

```
# Fitting to the data
M1 <- update(M1, prior_PD = F)
M2 <- update(M2, prior_PD = F)
M3 <- update(M3, prior_PD = F)
```

Leave one out cross validation (LOO) was used to compare the three models. This approuch tests model on unseen data and prevents overfitting. The model that maximized elpd provides the best fit to the data. Shown below in Table 1. Model 2 which includes social media experiance variables with an equal effect across different political ideaologies produced the best results.

Table 1: Model Comparisons

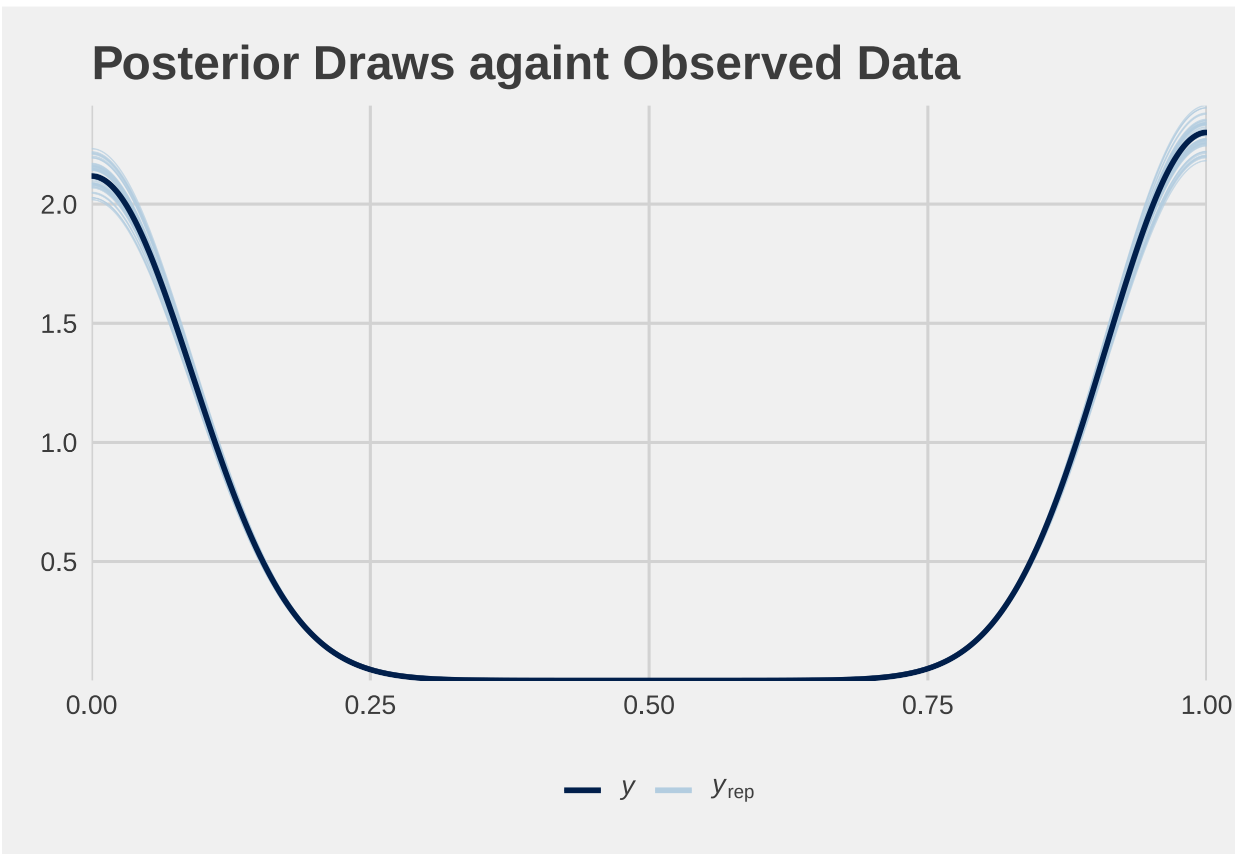
	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
M2	0.000000	0.000000	-2009.843	15.56040	15.56006	0.2620555	4019.687	31.12079
M3	-5.173901	2.235346	-2015.017	14.91949	23.52739	0.4172671	4030.035	29.83899
M1	-24.858873	7.573431	-2034.702	13.87719	11.87562	0.2280922	4069.405	27.75437

Diagnostics

The LOO comparisons shownn above indicated that model 2 was the best model relative to the other candiadte models, however, several daignostic tests are nessesary to ensure that the model is accuratly specifed. The LOO comparison shown above is a relarive measure and without proper diagnostic checks, it is possible that all 3 models were poorly fit. Assesment of various diagnositc tests suggests that the final model is well specified.

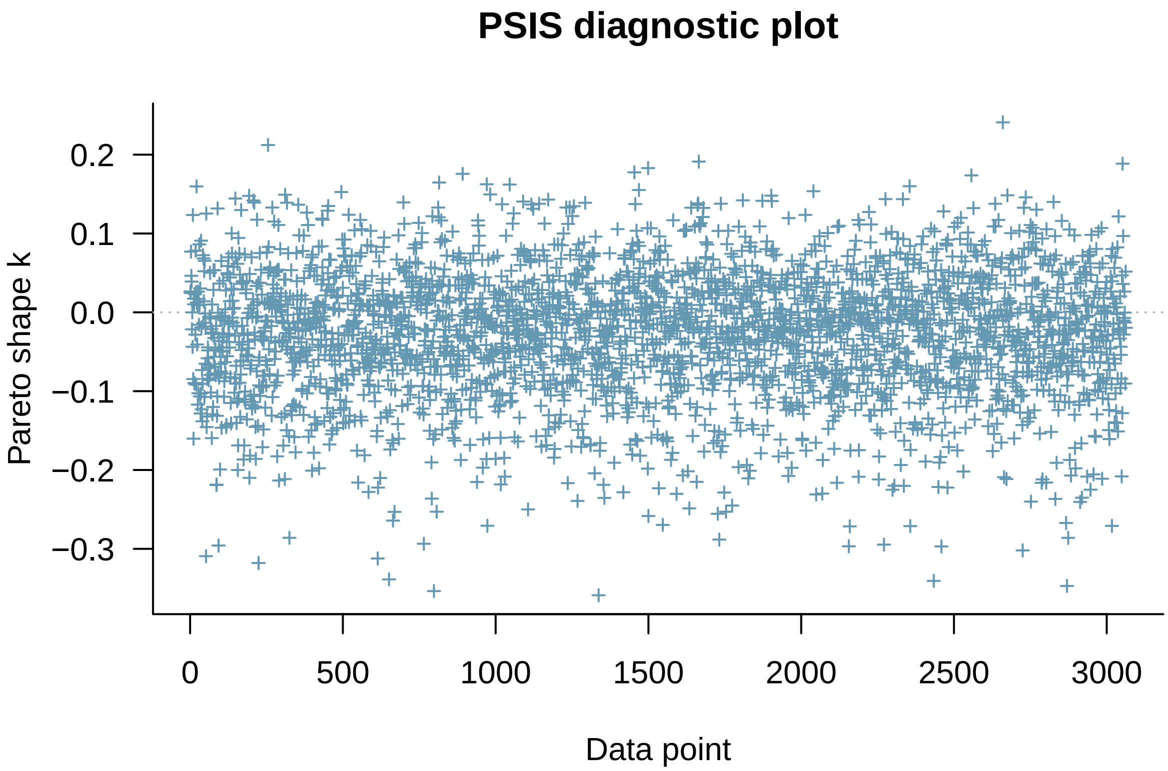
Posterior draws compared to the observed data

The plot below shows the observed data ploted over each draw from the posterior distribution. Results suggest the posterior distribution strongly fits the observed density of the actual data points and implies good model fit.



Checking for outliers and leverage points

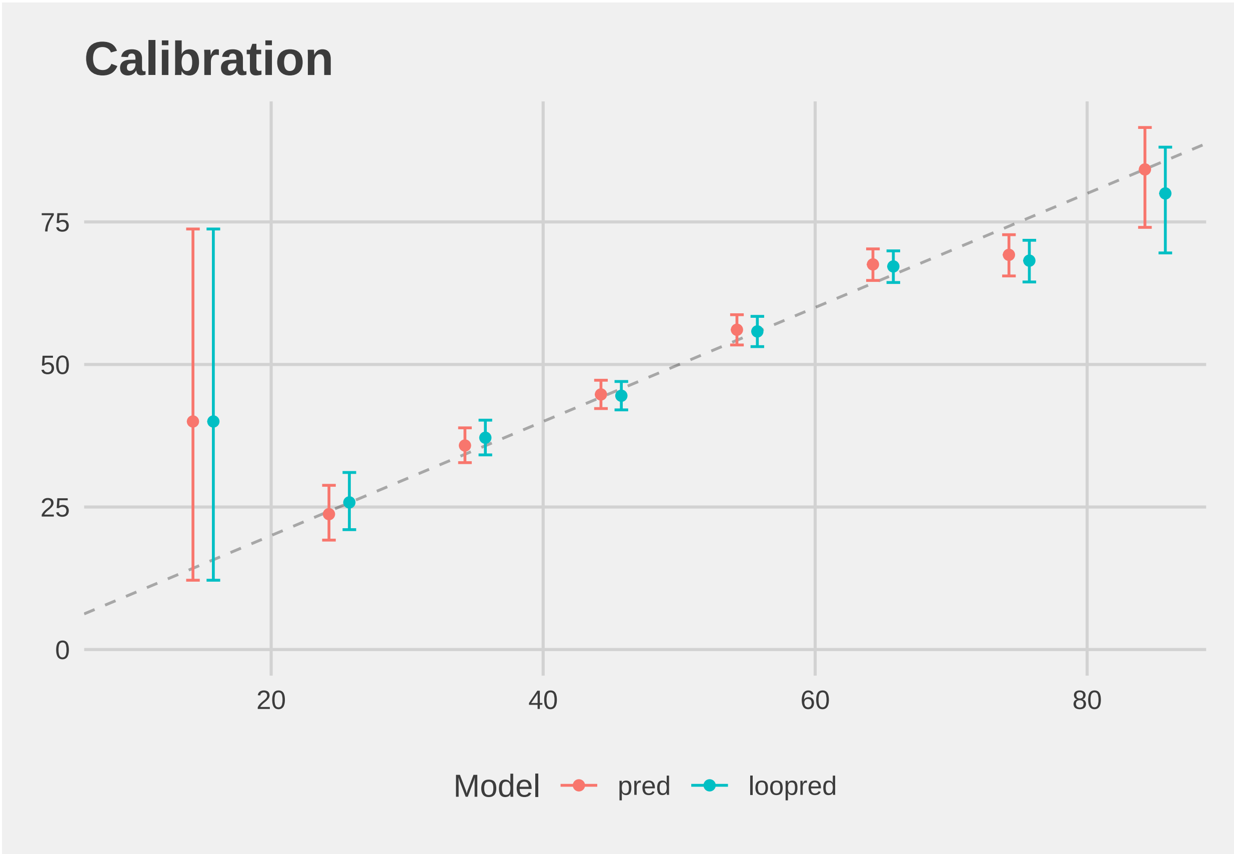
As with any general linear model, outliers and high leverage points can bais the regression line and are problematic. The plot below shows there are no outliers or high leverage points that bais the realizations of model paramterters.




```
## NULL
```

Calibration

Logisitc regression models return predictions in the form of percentages. The higher the percentage the more likely a data point is to be in class 1, the lower the percentage the more likly it is to be in class 0. Calibration plots compare predictor probabilties againt empiracle probabilites. When calibration is good, an event that is predicted to haven 70% time actually happens 70% of the time. In calibration plots, “good” calibration is reflected when points are close to the dignal line. Shown below, the final model is well calibrated and, in general, higher predicted probabilites coincide with higher empiriicle probabilities.



Inference: interpreting results

Model 2 had the best fit to the data which provides evidence that opinions about the regulation of social media comapines is not only a function of political identity but is related to the types of content users see while on social media sites. Moreover, Model 2 holds that the association between social media content and views on regulation is equal across different political ideaologies.

Parameters realizations

Table 2 presents the median value of each parameter along with a corespnding 90% intervals. \begin{table}

\caption{Model Estimates with 90% Credible Interval}

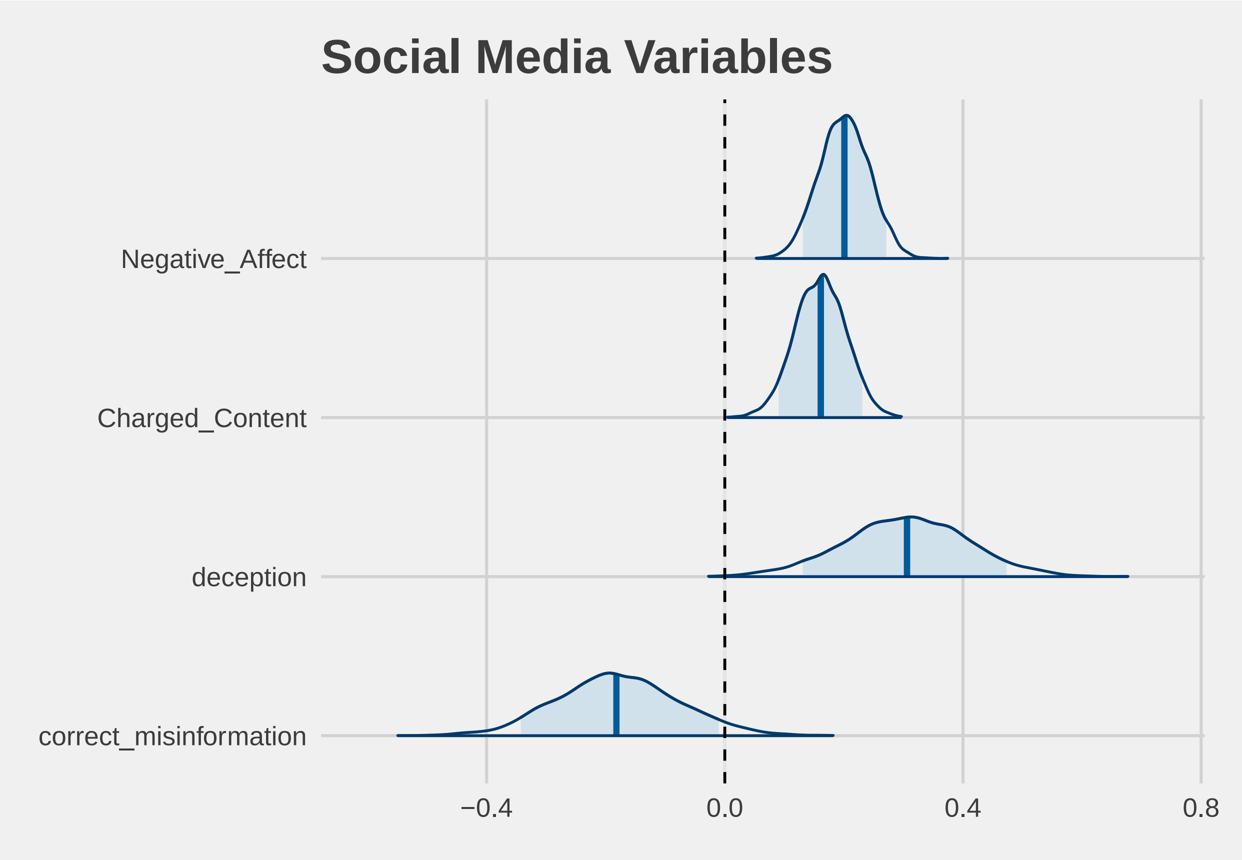
	Median	5%	95%
(Intercept)	-0.26	-0.41	-0.12
very_conservative	-0.63	-0.85	-0.40
conservative	-0.49	-0.65	-0.33
liberal	0.49	0.32	0.66
very_liberal	0.85	0.65	1.05
age_18_29	-0.04	-0.24	0.16
age_50_64	0.39	0.24	0.55
age_65_up	0.51	0.34	0.68
black	0.11	-0.11	0.33
asian	0.42	0.05	0.80
mixed_race	0.50	0.17	0.84
other	0.22	-0.01	0.45
Negative_Affect	0.20	0.13	0.27
Charged_Content	0.16	0.09	0.23
deception	0.31	0.13	0.47
correct_misinformation	-0.18	-0.34	-0.01

\end{table}

Social media content

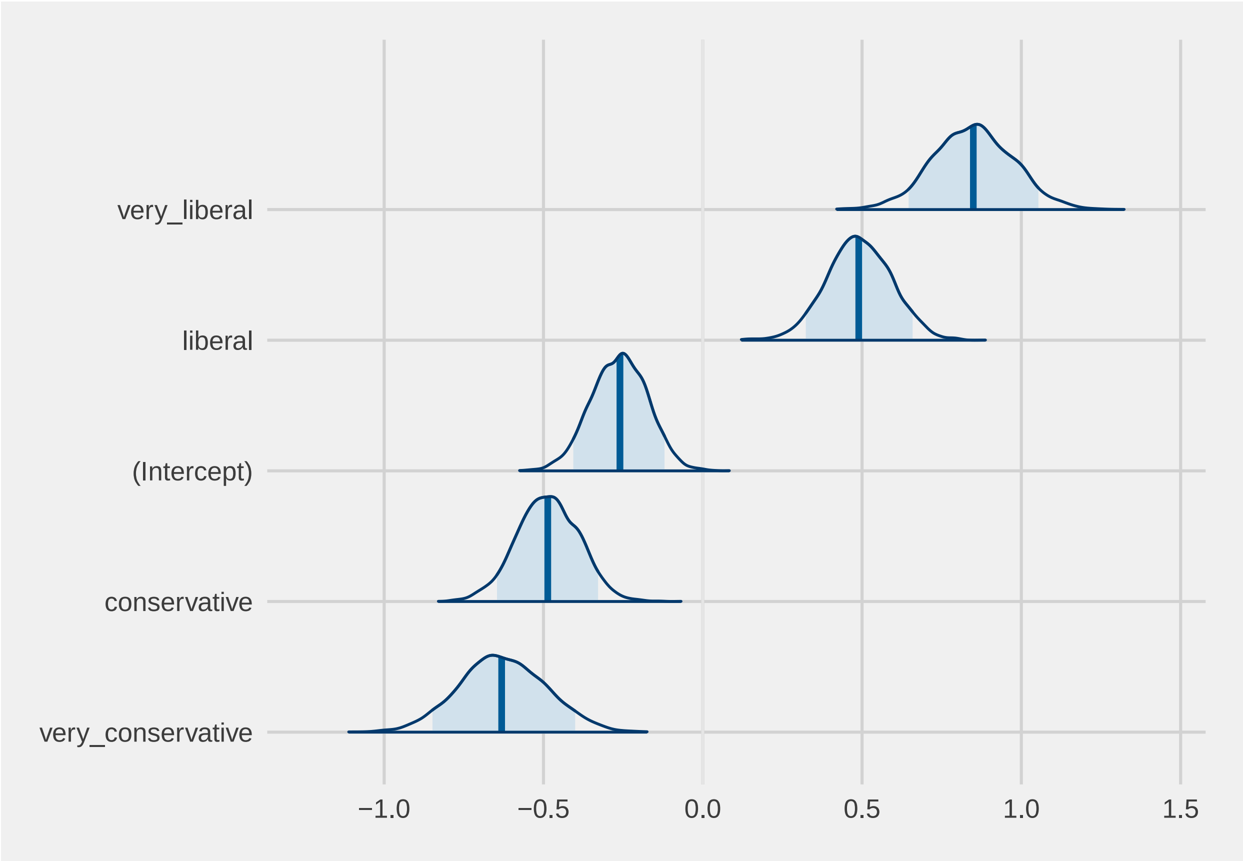
The plto below visulaizes varaibles related to experiance on social media sites. Holding political ideaology constant, users who see content that increases negative affect are more likely to favor increasing regulation. A similar patern emerged for charged content, where reports of seeing more politically charged and controversial posts was associated with favoring increased regulation.

There was more uncertanty surrunding the assocation of deception and correcting misinformation, this is reflected in the wider distributios. That being said, users who reported seeing more deceptive posts consistently favored increasing regulaton. The assocaiton between seeing more posts that try to correct misinformation was the most inconclussive of the social media variables, however, over 95% of draws from the posteriro distribution signaled that users who saw more posts attempting to correct misinformation were against expanding regulation.



Plotical ideaology

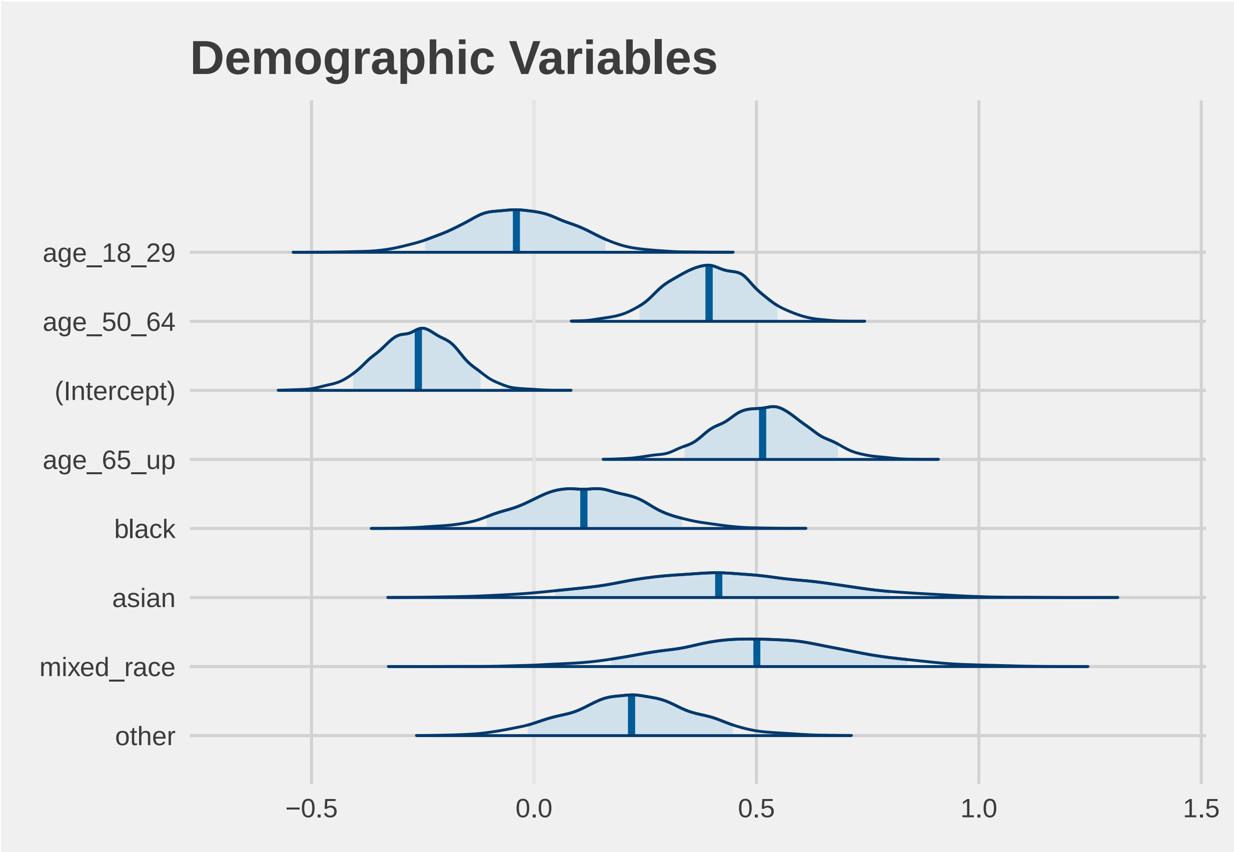
As expected there is a hgh degree of evidence realting political ideaology to views on regulation of social media. More liberal respondents consistently favored increasing regulations while more conservative respondents were against expanding regulations.



Other demographic identites

Understanding demogrpahic associations was not the main focus of the present analysis, however, several trends emerged among diffent age and rcial groups. Compared to younger age groups, those 50 (both ages 50-64 and 65 and up) were more in favor of expanding regulations. Compared to other races, whites were more agsint expanding regulations.

```
mcmc_areas(M2,
           pars =
             c("age_18_29",
               "age_50_64",
               "(Intercept)",
               "age_65_up",
               "black",
               "asian",
               "mixed_race",
               "other"),
           prob = .9) +
  theme_fivethirtyeight() +
  labs(title = "Demographic Variables")
```



Conclussion and Notes on Casueality

This analysis found that the types of content users report seeing is associated with their views on wehter or not social media compines need to be more regulated. This is valuable to social media compaines in that it identifes that downregulating deceptive informaiton and posts that generate negatvie affect may be an effective statagy to increase favorability among the general public. This analysis also idetifes several political, age and racial demogrpahics were social media compaines have an oppurtunity to improve realtions with.

The analysis conducted represent a stating point for more indepth studies, but it needs to be noted that the design used is not causal in nature. The results reported do not imply that social media content casues users to favor more or less regulation. While this is possible, it can not be confirmed throgh this study design. Subseqent intervention studies, should explore if downregulating posts with negative affect, charged content or disinformation direclty lead to changes in opinon.