Users' characteristics and online activity on Twitter: do they reflect a moral-emotional debate?

Vanessa Horacio Lira

Abstract

Is the use of moral-emotional words in Twitter comments about politics associated with user's characteristics and platform activity? Through a binomial logistic regression model it is expected to verify the association between the use of those words and the number of tweets, followers, likes, and profiles followed by a Twitter user. The characteristics will be measured as the presence (or not) of elements of religion and ideology, as well as gender. While controlled by issue and how long a user has that account. 9986 observations (tweets) were collected. The following sections are divided into literature review, theory, research design, and data, results, and conclusions.

Introduction

The relevance of the impact of social networks on political behavior has led to the increase of studies on the subject. Past studies have shown that the frequency at which people stay informed is crucial to explain how sensitive they are to the issue (Mouran, Urdinez & Onuki, 2016). At the same time, there is a greater concentration of political preferences as they do not connect to networks (SOUZA, GRAÇA & SILVA, 2017). Most importantly, the presence of disrespect in online comments (Mendonça & Amaral, 2016), and comments that radicalize the debate, using inappropriate and offensive terms (Mitozo, Massuchin & Carvalho, 2017).It is important to note that the platform is the main factor driving the provision of reasons on political comments (Mendonça & Amaral, 2016). This debate is important because it highlights an existing conflict for democracy. While conflicting opinions lead to more tolerance and greater understanding of different opinions, divergence discourages political participation and reduces the chances of correct voting (Bello, 2016).

The various studies on political behavior on social networks, at least two issues are key: the social networking platform matter and the content of shared comments has something to say. This article seeks to investigate whether user characteristics and platform activity are associated with the use of moral-emotional words in comments on

political issues. By moral-emotional words we refer to the sense of right and wrong that can affect the daily interactions of individuals, based on the theories of Psychology. Investigating how social networks convey moral norms and attitudes becomes critical (Brady et al., 2017).

Literature Review

The growth of informal policy discussions on social networks is a recent phenomenon that only grows around the world. But how common they are, how often they occur beyond party boundaries and whether they occur through strong ties or not are crucial issues in understanding whether online platforms are contributing to political polarization or just softening their effects. Who is more likely to talk about politics? Research points to a gender gap in such discussions. Men are more likely to say what they "discuss national politics almost every day" than women (31% to 20%) and they like political discussion more than women (36% to 26%) (Verba et al. , 1997).

Specifically, in the political interactions on Twitter it is possible to trace strong polarized retweet patterns, demonstrated by a high degree of clustering from the ideological bias of users. In other words, on political issues users interact positively within the same ideological bubble (Conover et al., 2011). From a comparative perspective, when we look at Twitter users' interaction on diverse topics, we find that salient policy topics resemble “echo chambers,” environments that potentiate ideological polarization, while other topics such as the Olympics or Super Bowl do. closer to “national conversations” with low level of polarization (Barberá et al., 2015).

The sharing of fake news on social networks is a phenomenon mainly associated with political discussions. Its occurrence on online platforms has already been debated in the literature, and there is no consensus on its impact on election results. However, research is still lacking in models that analyze the relationship between sharing fake news and its impact on individuals' trust in democratic institutions and, consequently, in the process of political polarization (Tucker et al., 2018).

In probabilistic terms the increase in political polarization was greater among individuals who do not use social networks compared with those who use these technologies to interact with others (Boxell et al., 2017). Therefore, Facebook and Twitter users are exposed to more diverse political opinions compared to those individuals who do not use social networks (Bakshy et al., 2015). The results of the survey conducted by

the Pew Research Center corroborate this finding, as most users reported that due to social networks they are exposed to a wide range of policy views (Duggan and Smith, 2016).

Twitter itself presents itself as the platform where people seek to interact about politics, studies have shown that political interactions between users with different political ideologies are more frequent than expected (Barberá et al., 2015). However, Barnidge (2017) argues that individuals have more misunderstandings in interactions on social networks than in person. This shows how the literature proposes more questions than it answers (Tucker et al., 2018).

Research on the subject has so far found conflicting results on the effects of social networks on political polarization. Fletcher and Nielsen (2017) conclude that people who use social networks are exposed to various news at a higher level than people who do not. Understanding why most political information shared on social networks is partisan or extremist can be made because users differ in how active they are on social networks as well as in content production.

Twitter users with more extreme ideological positions share disproportionately more content than moderate users (Barberá and Rivero, 2015); Preotiuc-Pietro et al., 2017). There is a fundamental imbalance in sharing information on Twitter, because even though most users post links to ideologically moderated news sources, a small group of users share ideologically extreme content and can be responsible for the majority of published policy tweets (Shore et al., apud Tucker et al., 2018).

In addition, we can highlight two factors about political polarization. First, there is a consensus that elite behavior is driving political polarization. That said, messages that emphasize party conflict reinforce polarization, while messages that emphasize intra-party conflict have the potential to reduce it. Second, emotions are important, because while anger makes people less likely to be suspicious of inaccurate information that supports their views and more likely to distribute it; Anxiety may have the opposite effect, prompting individuals to seek accuracy rather than directional goals (Tucker et al., 2018).

Theory

Based on the results found in the briefly synthesized literature above, the article proposes to identify whether the use of moral-emotional words in political tweets is associated with the activities and characteristics of users. It is expected to find some association between the number of tweets, followers, likes, and profiles followed by a Twitter user and the use of those words. The characteristics will be measured as the presence (or not) of elements of religion and ideology, as well as gender. While controlled

by issue and how long a user has that account. It is expected that users with profiles with elements of ideology and/or religion to be more likely to use moral-emotional words, as well as the newer users on the platform.

The choice of Twitter social network is due to two factors: first, the availability of data that this platform offers - we collected thirty-one variables; second, Twitter being the most used platform by current president Jair Bolsonaro for communication with the public since the election campaign. Therefore, it is expected that more people will use the platform and thus it will become an increasingly representative sample of the population.

The research differs due to: 1) the choice of the unit of analysis, that is, tweets from ordinary individuals, not representatives or influential people; 2) time period analyzed: from specific contexts (Copa America and Social Security Reform), not in electoral periods, in which we have an isolated event and an event on a politically polarized issue; 3) inclusion of variables from literature review on the topic. Assumptions:

H1: The use of moral-emotional words is associated with the individual's activity on Twitter.

H2: The use of moral-emotional words is associated with the characteristics of the individual on Twitter.

H3: There is greater use of moral-emotional words during the welfare reform vote.

H4: Newer Twitter accounts use more moral-emotional words.

H null: There is no association between the activities and characteristics of individuals and the use of moral-emotional words.

Research Design and Data

As the purpose of the article is to identify the behavior of individuals in the social network, we chose two distinct moments for data collection, but at the moment when these events were happening (live broadcast). The first event is President Jair Bolsonaro's television appearance at the 2019 Copa America (07/07/2019). The second event was the

voting period in the House of Representatives on pension reform (07/10/2019). Although events differ in time of data collection, they have a similar number of observations. While in the first database we had a total of 24,564 observations, in the second we had a total of 25,616. After pre-processing the data, we have 9.986 observations. The pre-processing data excluded tweets based on three criteria: those not written in Portuguese, the retweets, and the repeated tweets.

Data collection was made through Twitter API platform that provides data from searching tweets. 31 variables were collected. The dependent variable was measured using the text variable. It was classified as binary, given the presence or absence of the use of moral-emotional words. The list of words was the same as that applied in Brady's article, but with the proper translation by the author. In the end the database had a discrepancy in the data of the dependent variable. That is, the proportion of events was much smaller than the proportion of non-events. Therefore, there was a bias in the variable. Since ideally the ratio of events and non-events in Y should be roughly the same, we sampled the observations in approximately equal proportions for better models.

Through a logistic regression model, we expect to check whether the use of moral-emotional words in political tweets is associated with the activities and characteristics of users. For this, the independent variables will consist of the users' activities, measured through the variables: number of tweets, followers, likes, and profiles followed by a Twitter user. And the user's characteristics, measured as the presence (or not) of elements of religion and ideology, as well as gender. Control variables include event and user profile creation date. The models can be summarized in the equations below follow by a Table 1 with the list of variables:

Model 1

$$\text{Moral/emotional words} = \alpha + \beta1(\text{posts}) + \beta2(\text{followers}) + \beta3(\text{likes}) + \beta4(\text{followed}) + \mu$$

Model 2

$$\text{Moral/emotional words} = \alpha + \beta1(\text{posts}) + \beta2(\text{followers}) + \beta3(\text{likes}) + \beta4(\text{followed}) + \beta5(\text{issue}) + \beta6(\text{time}) + \mu$$

Model 3

$$\text{Moral/emotional words} = \alpha + \beta1(\text{posts}) + \beta2(\text{followers}) + \beta3(\text{likes}) + \beta4(\text{followed}) + \beta5(\text{ideology}) + \mu$$

Model 4

Moral/emotional words = α + β1(posts) + β2(followers) + β3(likes) + β4(followed) + β5(religion) + μ

Model 5

Moral/emotional words = α + β1(posts) + β2(followers) + β3(likes) + β4(followed) + β5(gender) + μ

Model 6

Moral/emotional words = α + β1(posts) + β2(followers) + β3(likes) + β4(followed) + β5(ideology) + β6(religion)+ β7(gender)+ β8(issue)+ β9(time) + μ

Model 7

Moral/emotional words = α + β1(posts) + β2(issue)+ β3(gender) + μ

Table 1. List of variables

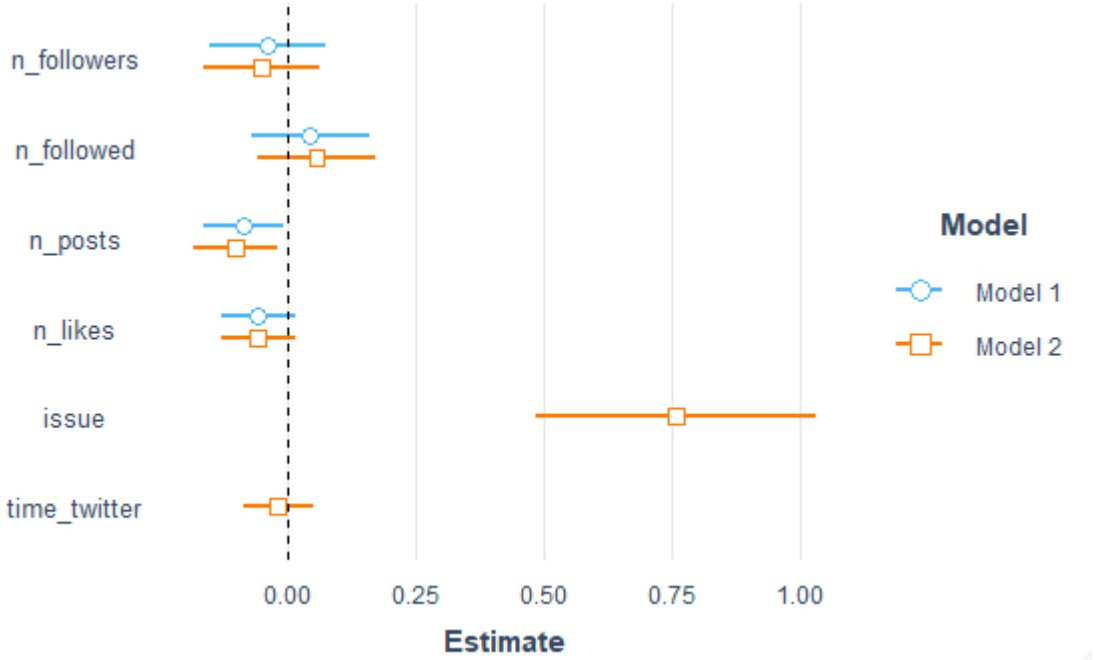
Variables	**Type**	**What is**	**Collect**	**Model**
Use of moral-emotional words	Binary (0 = No, 1 = Yes)	Sellection of moral-emotional words listed in Brady et al. (2017)	Transformed variable. Text Analysis.	Dependent variable
Number of posts (tweets)	Numeric	The number of Tweets this user has liked in the account's lifetime.	Twitter API	Independent variable
Number os followers	Numeric	The number of followers this account currently has.	Twitter API	Independent variable
Number of likes	Numeric	The number of Tweets (including retweets) issued by the	Twitter API	Independent variable

		user.		
Number of profiles followed	Numeric	The number of users this account is following.	Twitter API	Independent variable
Ideology	Binary (1= No; 2 = Yes)	Presence (or not) of ideological words in user's profile description.	Transformed variables through the variable *description*	Independent variable
Religion	Binary (0 = No; 1 = Yes)	Presence (or not) of religious words in user's profile description.	Transformed variables through the variable *description*	Independent variable
Gender	Binary (0 = non-identified 1 = Female, 2 = Male)	Gender of users based on their names.	Transformed variable through the variable *name.* GenderBR package	Independent variable
Issue (event)	Binary (0 = Copa America; 1 = Pension Reform	Tweet published in one of the events.	Classified by the author	Control variable
Time of Twitter	Numeric	Number of years a user has an account on Twitter	Twitter API. Transformed variables through the variable *user created at*.	Control variable

Findings and Results

In this section we will present the main findings and results of our logistic return models. In each section one of the five hypotheses of our study will be evaluated. We use the comparison between different models to increase our ability to assess hypotheses.

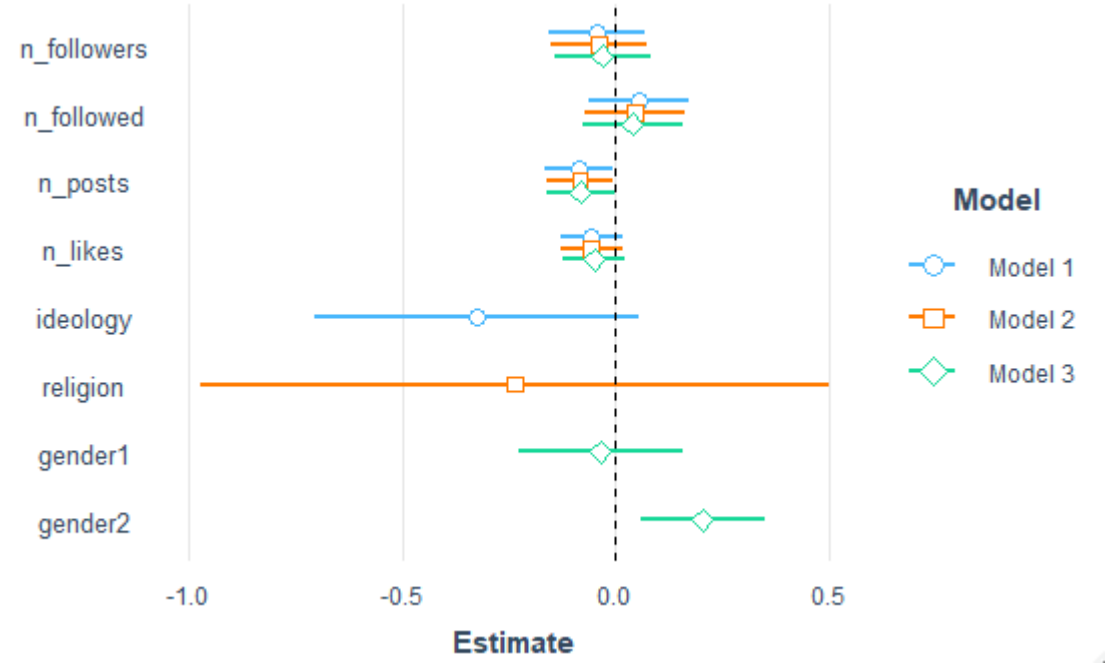
Hypothesis 1: Comparing models 1 and 2



The graph represents the estimated coefficients in the models for each variable over a 95% confidence interval. Although in the logistic regression we have named the models from 1 to 7, in graphics we will see in the legend the models named as: Model 1, Model 2, and Model 3. Thus, it is important to note which models the graphics are referring to only the coefficients of the variable "number of posts" do not touch 0 for both models. That is, even if it has a small effect, it is possible to say that the number of posts has a negative association with the likelihood of using moral-emotional words.

Similarly, for model 2 the issue variable does not touch 0. In this case, having a higher level of association comparatively and positively with respect to the variable number of posts. As our hypothesis 1 seeks to measure the association between the four variables that compose user activity on Twitter with the use of words, it is possible to say from the analysis of the models that it is more likely that there is no association between users' activities and the use of moral-emotional words. Importantly, the use of controls improves the fit of the model, but with very little impact. We see that the addition of variables makes the number of posts a bit far from zero. But practically, the addition of variables does not make changes in the coefficients estimates.

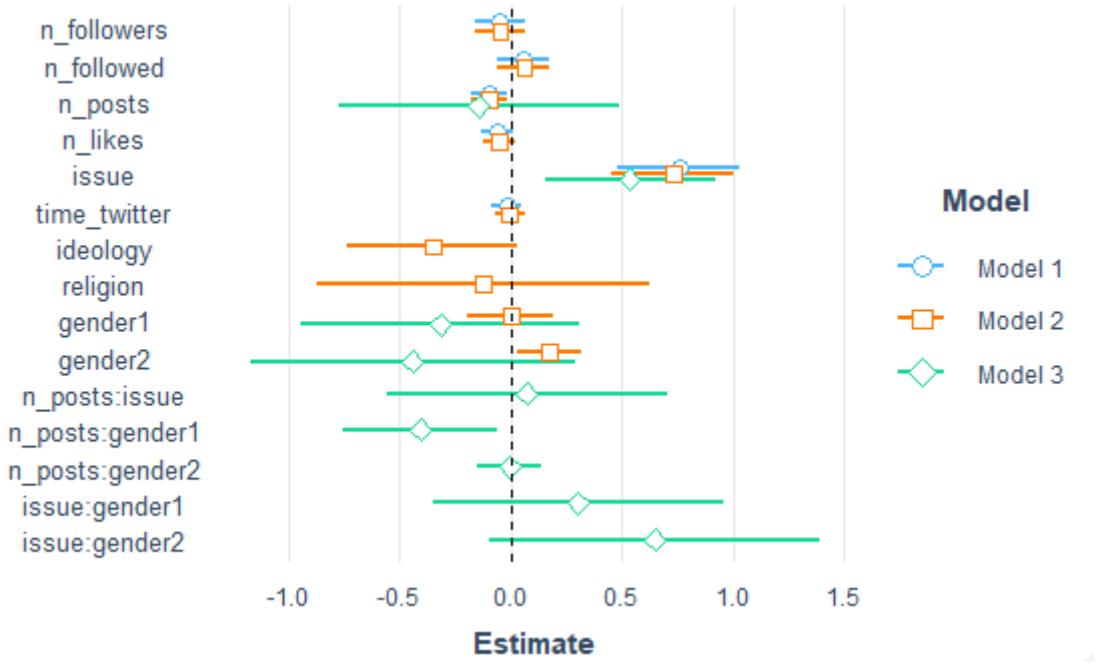
Hypothesis 2: Comparing models 3, 4, 5



When comparing the models concerning the hypothesis 2, we observe that the variable number of posts is the only one that does not touch zero, although it is negative. This means that this variable might have a negative association with the dependent variable (use of moral-emotional words), but it has a small effect, given its closeness to zero.

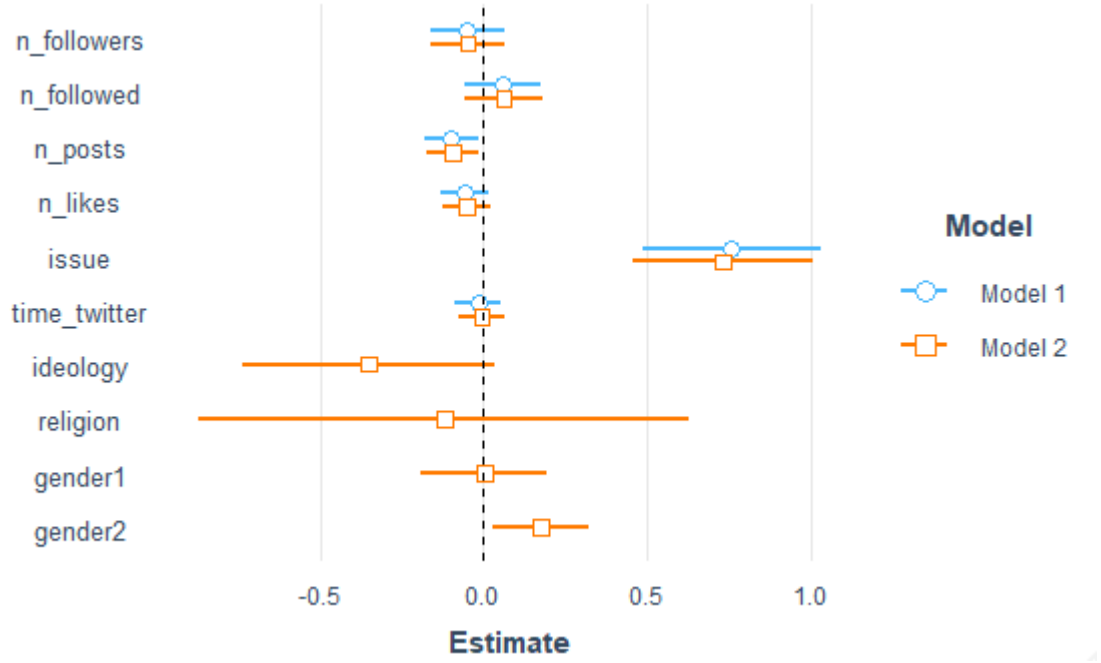
As we can see in the graphic the variables ideology and religion present coefficients with higher confidence intervals. Another relevant variable is gender2, presented in model 5. It does not touch zero and its coefficient is positive. This means the if the user is a man it is more likely that he uses moral-emotional words on political tweets.

Hypothesis 3: Comparing models 2, 6, and 7



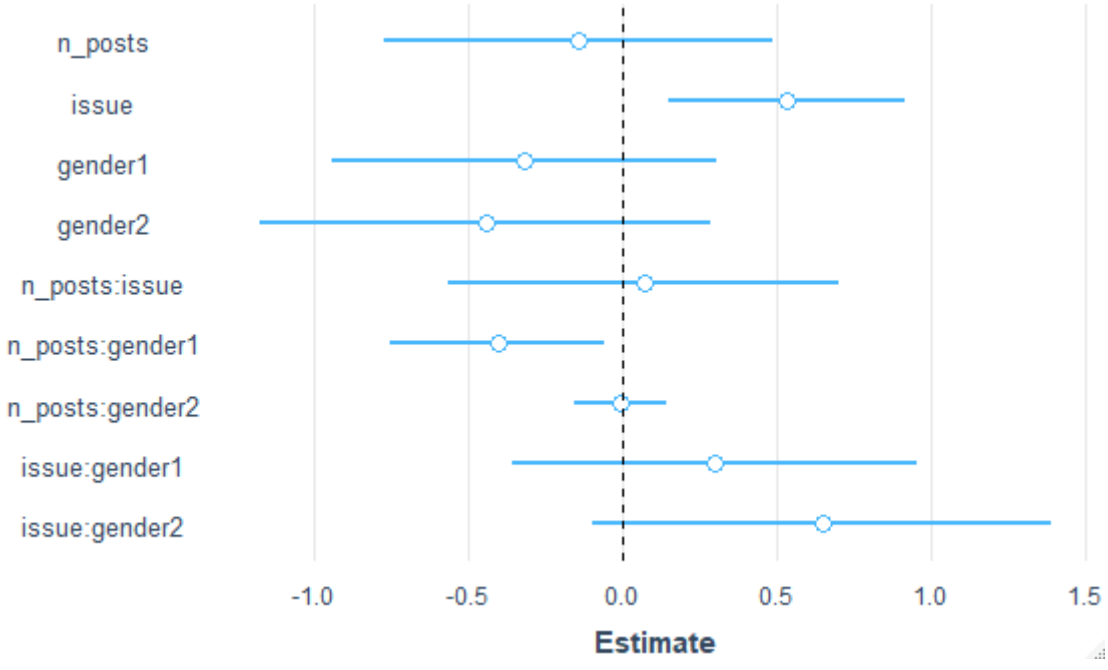
We observe that the coefficient of variable number of posts does not touch 0, except in model 7. The variable issue is positive in all models, and does not touch zero, which means that is relevant to predict that the tweets published on the issue, probably pension reform, are more likely to use moral-emotional words. Also, in model 6, being a man is positively associated with the use of moral-emotional words on political tweets. And being a woman and have higher number of posts is negatively associated with the use of words.

Hypothesis 4: Comparing models 2 and 6



The hypothesis 4 expects that newer Twitter accounts use more moral-emotional words. As we can see from the graphic, the confidence interval from the time of Twitter coefficients touches zero. Therefore, it is very unlikely that exists association with the dependent variable. In other words, we have no evidence that users with more time of Twitter are more or less likely to use moral-emotional words on political tweets.

Hypothesis 5: Model 7



Is the use of moral-emotional words associated with the interactive relation between the variables gender, number of posts, and issue? These variables are the ones that did not touched zero in the other models. Nevertheless, they do not have explanatory power, because only two of them do not touch zero: issue, with a positive effect, and number of posts from women, which has a negative effect. This means that the higher the number of posts from female users, the lowest is the probability of using moral-emotional words. However, in both the effects are negative limiting the explanatory capacity of the model.

In short, the analysis of the models shows us mixed results. While some variables, such as number of posts, gender, and issue have some effect, others have several problems and limitations, as shown above in the confidence intervals of the estimated coefficients. With this, we note that it is, although we cannot immediately reject the null hypothesis, but it is more likely to be true. Evidence indicates that the association between the selected independent variables is weak or does not exist. Therefore, they may not be the best predictors of whether or not an individual commenting on politics on Twitter will use moral-emotional words. It may also be necessary to repeat the models based on other random samples.

Implications and Conclusions

The impact of social networks on electoral disputes is still an expanding topic in the study of political science. However, there is debate about the association between social networks and political polarization in various countries of the world. In our literature review we showed how some explanatory variables were identified in preliminary studies on the topic. We can highlight the level of activity of users, their profile, time on social network and topic discussed. Another key point is the importance of twitter as a tool for global political interaction. In which several state leaders have personal profiles in which they dialogue with opponents and citizens, Donald Trump and Jair Bolsonaro are examples.

The research proposed to analyze the relationship of these variables and the polarized behavior on twitter. Measuring ideological bias from people's tweets is a major methodological challenge. We used as a rough measure the model of moral-emotional words of Brady et al. (2017). Thus, the use of these terms was considered as a polarized

interaction. This measure is an important limitation of our model. When testing our hypotheses from seven logistic models we find little evidence that the variables mentioned in the literature are associated with the occurrence or not of polarized interaction. However, some factors such as number of posts, genre and topic of discussion can have some effect even when controlled by other variables. In other words, the results found favor the null hypothesis, however it would be hasty to exclude the other hypotheses in a hasty way.

In short, our models find little explanatory capacity. Most of the model coefficients find values close to or equal to zero. Therefore, it would be crucial to replicate these models in larger random samples, so perhaps we would find more robust results. However, our study contributes to the literature mainly in questioning the real validity of the explanatory variables proposed in the current debate. Perhaps our results demonstrate the need to look for new explanatory alternatives to ideological polarization on twitter.

References

BAKSHY, Eytan; MESSING, Solomon; ADAMIC, Lada A. Exposure to ideologically diverse news and opinion on Facebook. Science, v. 348, n. 6239, p. 1130-1132, 2015.

BARBERÁ, Pablo et al. Tuíteing from left to right: Is online political communication more than an echo chamber?. Psychological science, v. 26, n. 10, p. 1531-1542, 2015.

BARBERÁ, Pablo; RIVERO, Gonzalo. Understanding the political representativeness of Twitter users. Social Science Computer Review, v. 33, n. 6, p. 712-729, 2015.

BARNIDGE, Matthew. Exposure to political disagreement in social media versus face-to-face and anonymous online settings. Political Communication, v. 34, n. 2, p. 302-321, 2017.

BELLO, André. A lógica social do voto correto no Brasil. Opinião Pública, v. 22, n.2, p. 466-491, 2016.

BESSI, Alessandro; FERRARA, Emilio. Social Bots Distort the 2016 U.S. Presidential Election Online Discussion. First Monday, v. 21, n. 11, 2016.

BOXELL, Levi; GENTZKOW, Matthew; SHAPIRO, Jesse M. Is the internet causing political polarization? Evidence from demographics. National Bureau of Economic Research, 2017.

Brady, William, Julian Willis, John T. Jost, Joshua A. Tucker, and Jay Van Bavel. (2017). "Emotion Shapes Diffusion of Moral Content in Social Networks." Proceedings of the National Academy of Sciences. 114(28): 7313-7318.

CONOVER, Michael D. et al. Political polarization on twitter. In: Fifth international AAAI conference on weblogs and social media. 2011.

FLETCHER, Richard; NIELSEN, Rasmus Kleis. Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication. Journal of Communication, v. 67, n. 4, p. 476-498, 2017.

FORELLE, Michelle et al. Political bots and the manipulation of public opinion in Venezuela. 2015.

Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2015). The structural virality of online diffusion. Management Science, 62(1), 180-196

IMAI, Kosuke. Quantitative social science: An introduction. Princeton University Press, 2017.

KOLLANYI, Bence; HOWARD, Philip N.; WOOLLEY, Samuel C. Bots and automation over Twitter during the US election. Data Memo. Oxford, UK: Project on Computational Propaganda, v. 2016, n. 4, p. 1-5, 2016.

MENDONÇA, Ricardo Fabrino; AMARAL, Ernesto FL. Racionalidade online: provimento de razões em discussões virtuais. Opinião Pública, v. 22, n. 2, p. 418-445, 2016.

MITOZO, Isabele Batista; MASSUCHIN, Michele Goulart; CARVALHO, Fernanda Cavassana de. Debate político-eleitoral no Facebook: os comentários do público em posts jornalísticos na eleição presidencial de 2014. Opinião Pública, v. 23, n. 2, p. 459-484, 2017.

MOURON, Fernando; URDINEZ, Francisco; ONUKI, Janina. Framing effects on foreign policy: experimental evidence from emerging countries and the Argentine-Brazilian rivalry. Opinião Pública, v. 22, n. 1, p. 195-218, 2016.

PHILLIPS, Whitney. This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture. Mit Press, 2015.

PREOŢIUC-PIETRO, Daniel et al. Beyond binary labels: political ideology prediction of twitter users. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017. p. 729-740.

Sanovich, Sergey, Denis Stukal, and Joshua A. Tucker. (2018). “Turning the Virtual Tables: Government Strategies for Addressing Online Opposition with an Application to Russia.” Forthcoming in Comparative Politics.

SOUZA, Rafael Martins de; GRAÇA, Luís Felipe Guedes da; SILVA, Ralph dos Santos. Politics on the web: using Twitter to estimate the ideological positions of Brazilian representatives.Brazilian Political Science Review, v. 11, n. 3, p. 1-26, 2017.

TRERÉ, Emiliano. (2016). "The dark side of digital politics: Understanding the algorithmic manufacturing of consent and the hindering of online dissidence." IDS Bulletin 47.1.

TUCKER, Joshua A. et al. Social media, political polarization, and political disinformation: A review of the scientific literature. Political Polarization, and Political Disinformation: A Review of the Scientific Literature (March 19, 2018), 2018.

VERBA, Sidney; BURNS, Nancy; SCHLOZMAN, Kay Lehman. Knowing and caring about politics: Gender and political engagement. The journal of politics, v. 59, n. 4, p. 1051-1072, 1997.

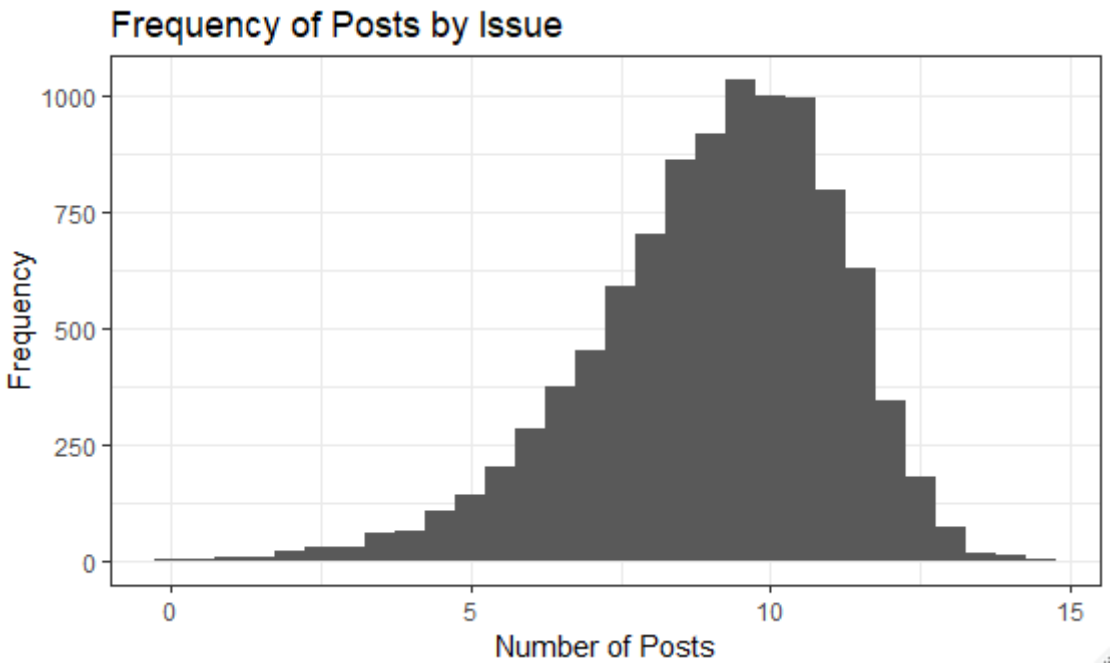
WOOLLEY, Samuel C.; HOWARD, Philip N. Computational propaganda worldwide: Executive summary. Computational Propaganda Research Project, p. 2017-11, 2017.

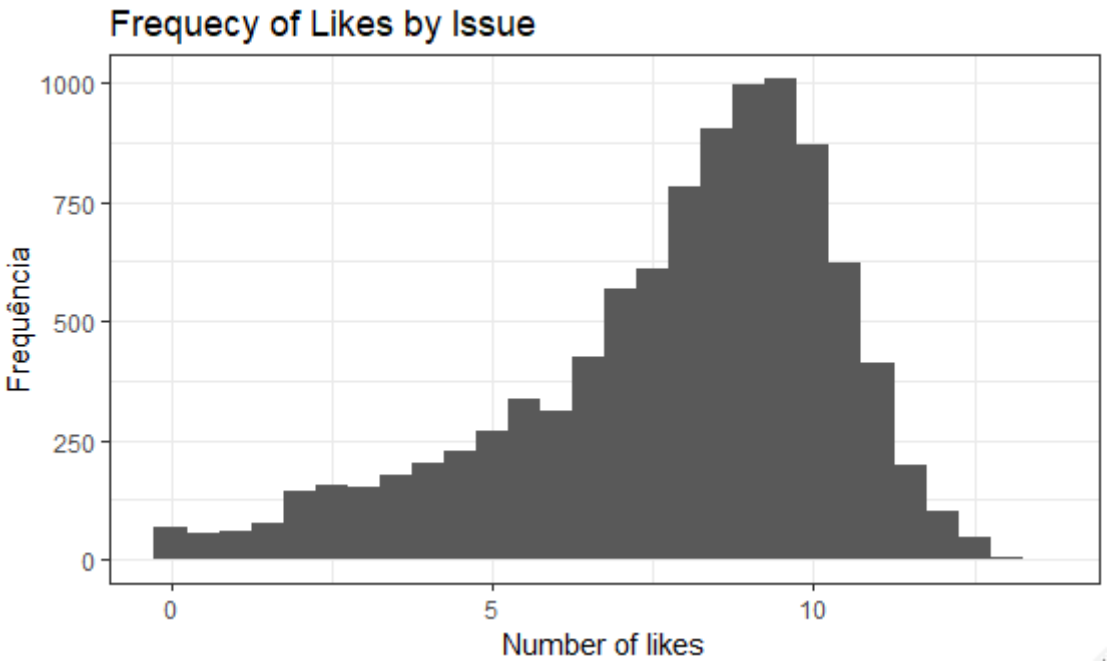
WOOLLEY, Samuel C. Automating power: Social bot interference in global politics. First Monday, v. 21, n. 4, 2016.

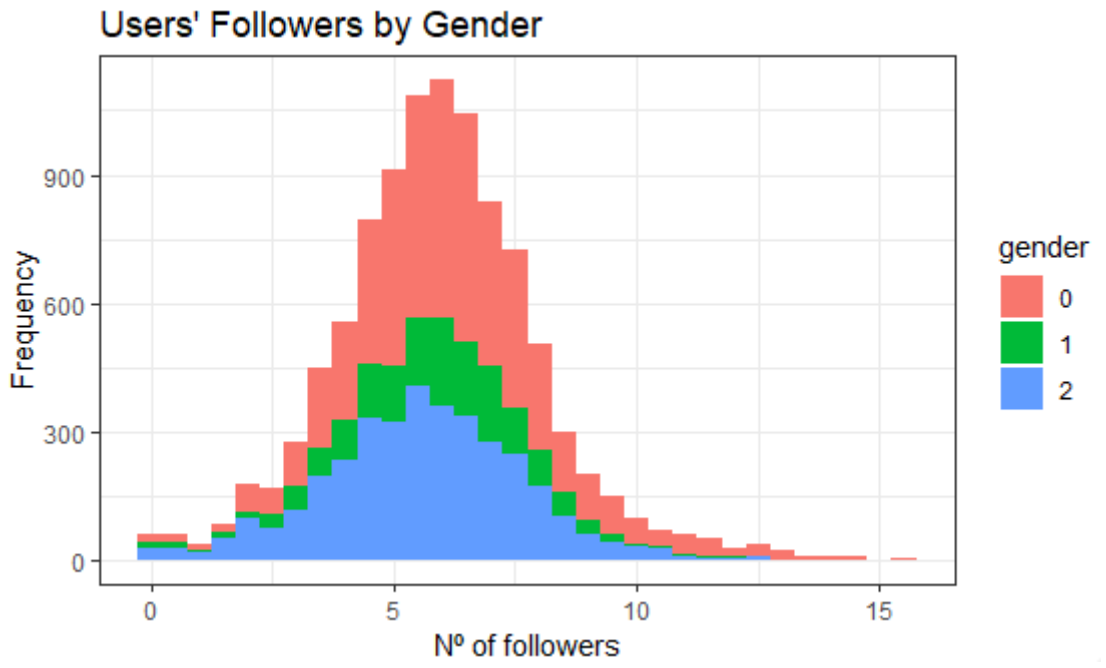
Appendix

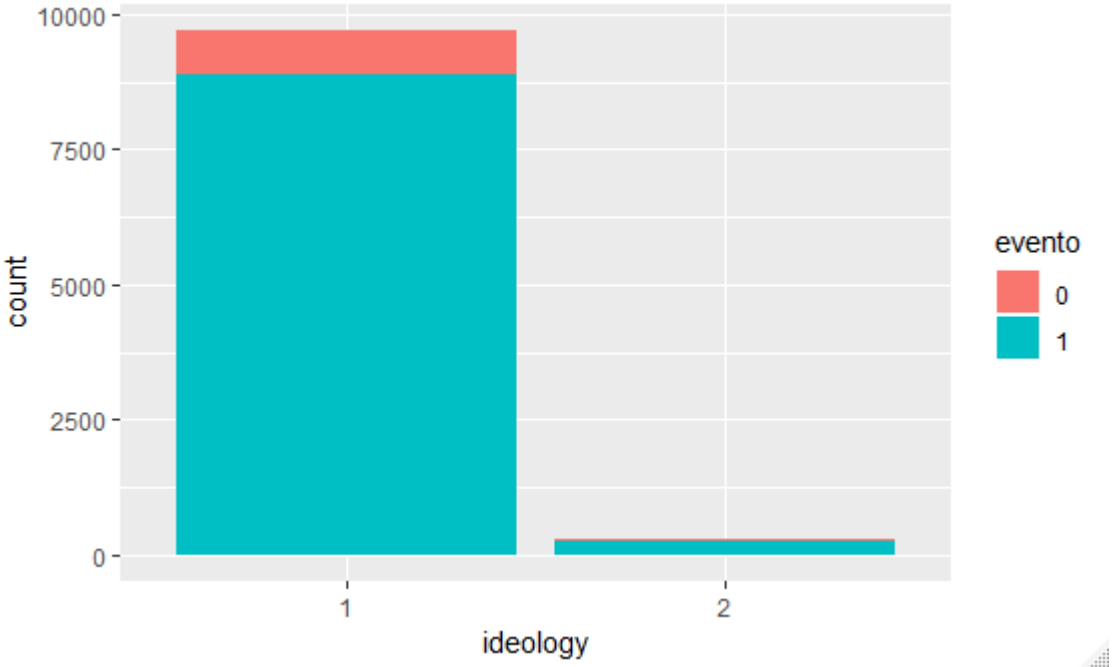
Exploratory Data Analysis

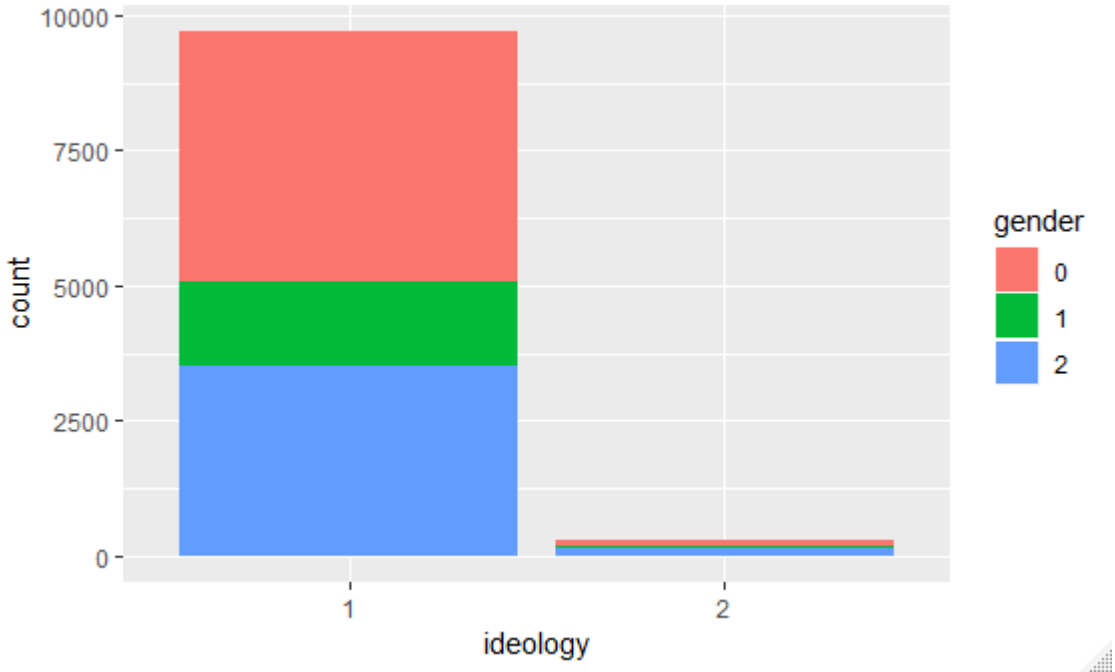
In this section we will use graphical analysis to visualize data distribution and check relationships between variables. We will see first the analysis on numerical variables, then the categorical variables, and finally the combination of both.

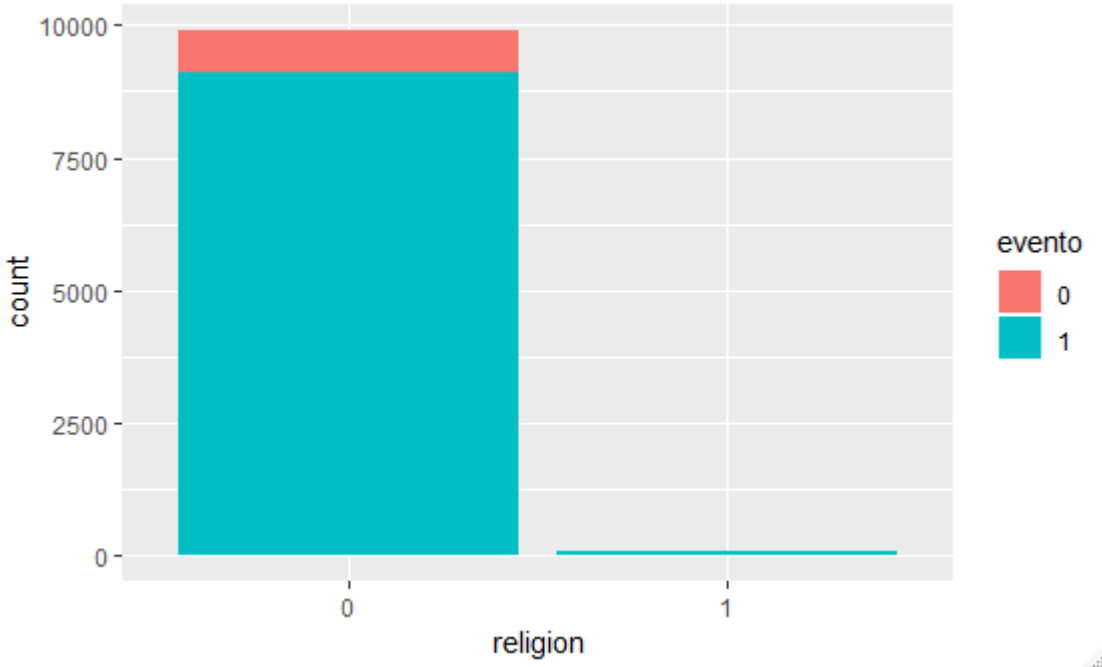


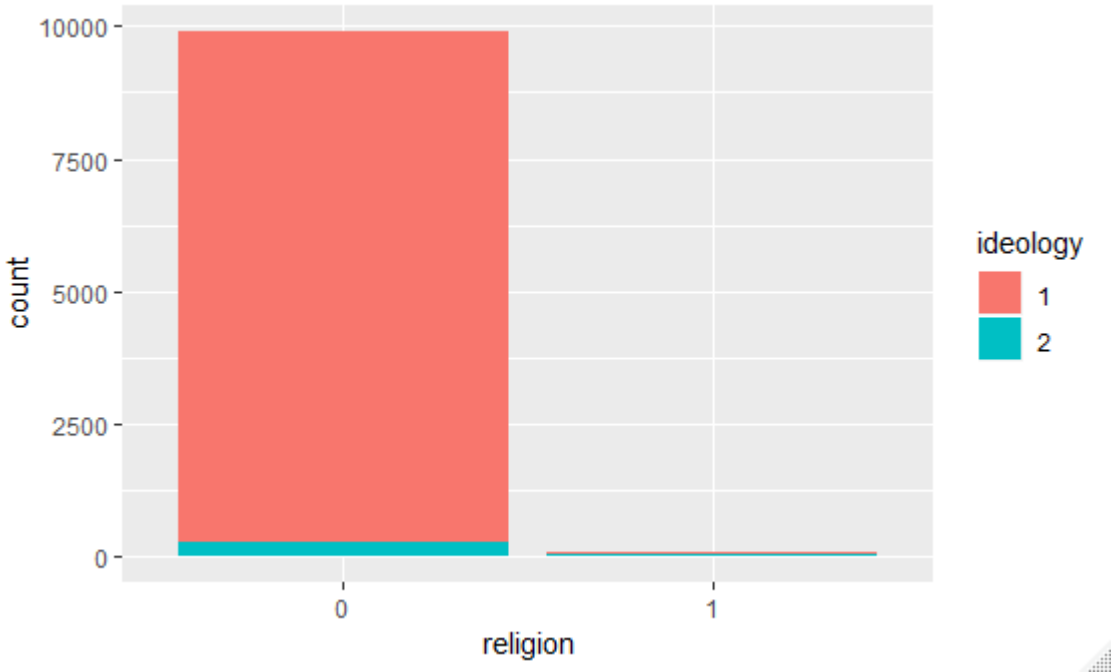

Frequecy of Likes by Issue
1000
750
500
250
0
Frequência
0
5
10
Number of likes

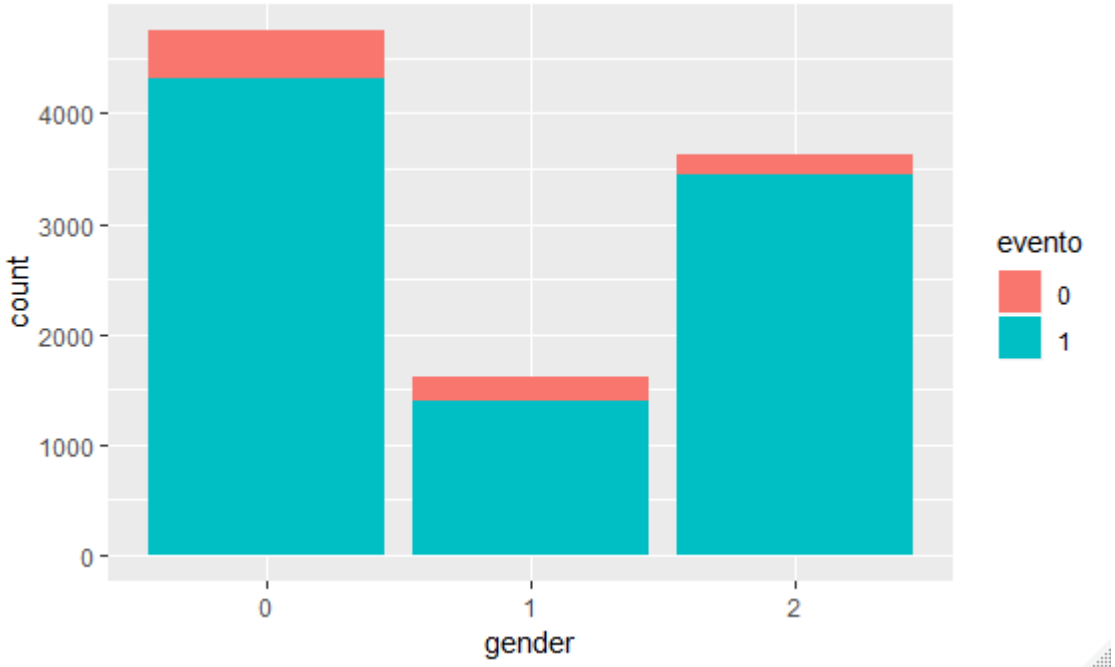

Users' Followers by Gender
900
600
300
0
Frequency
0
5
10
15
Nº of followers
gender
0
1
2

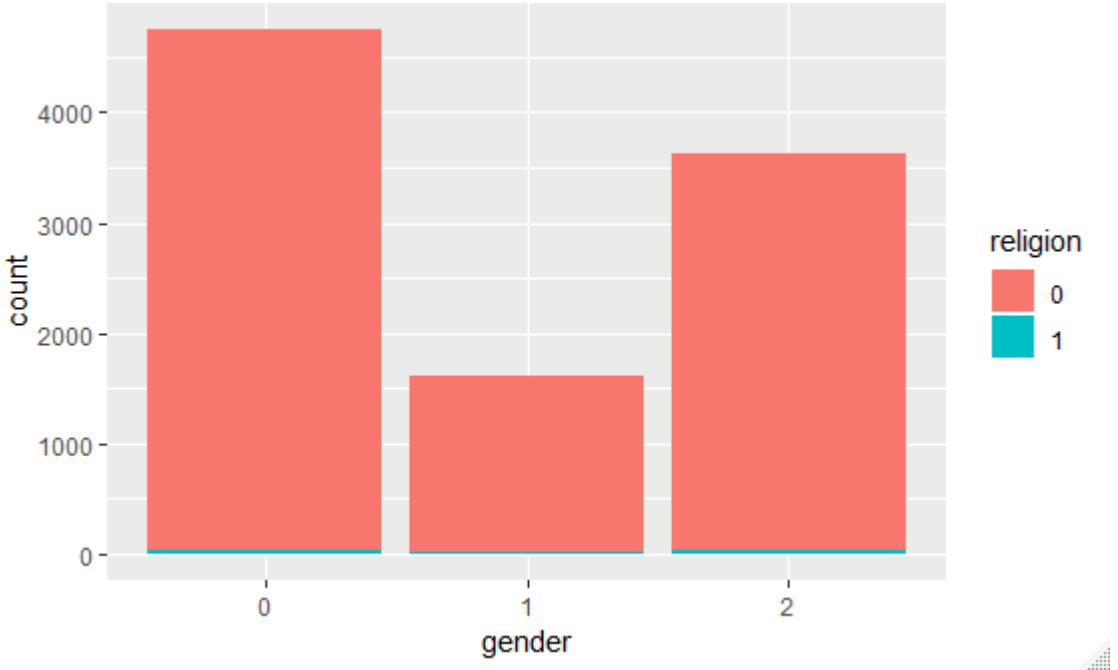

10000
7500
5000
2500
0
count
1
2
ideology
evento
0
1

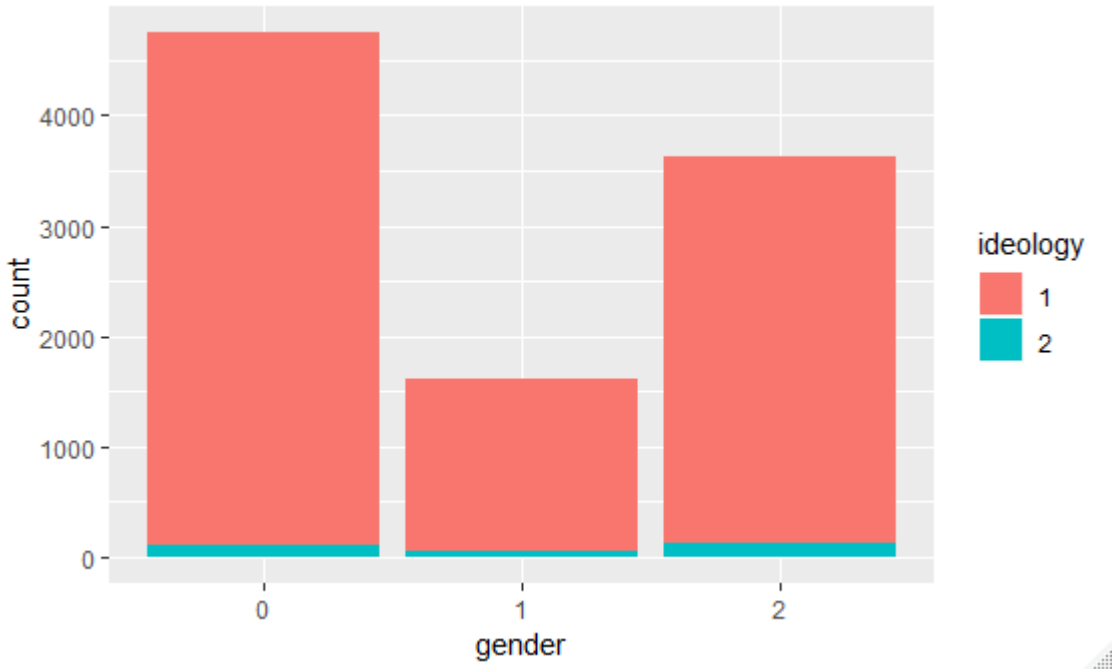

10000
7500
5000
2500
0
count
1
2
ideology
gender
0
1
2


10000
7500
5000
2500
0
count
0
1
religion
evento
0
1

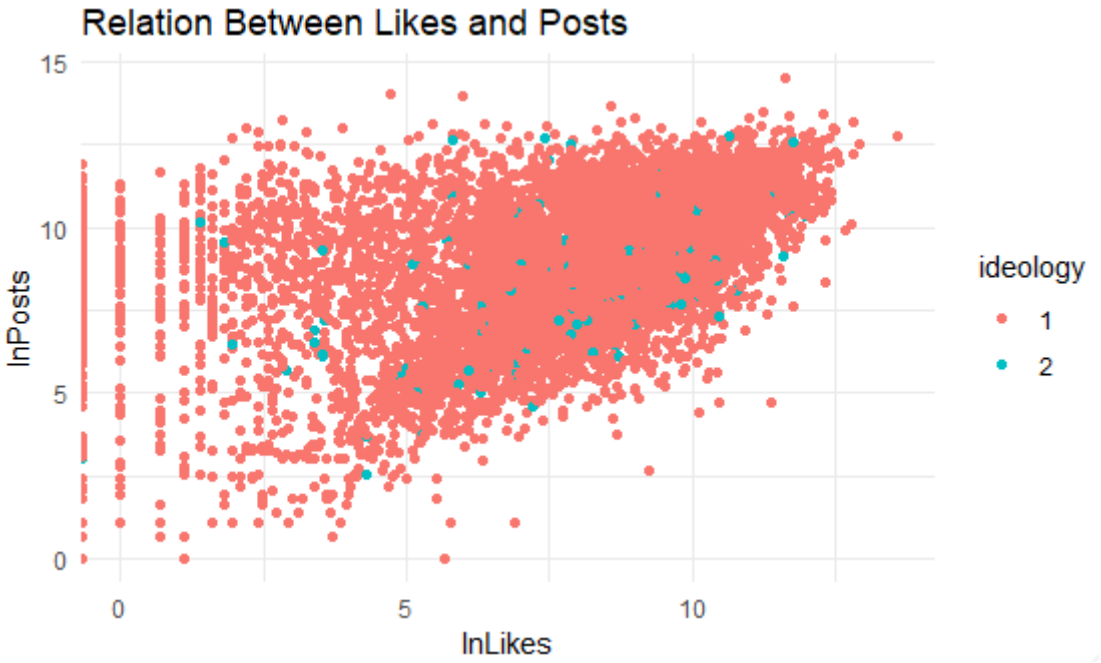

10000
7500
5000
2500
0
count
0
1
religion
ideology
1
2

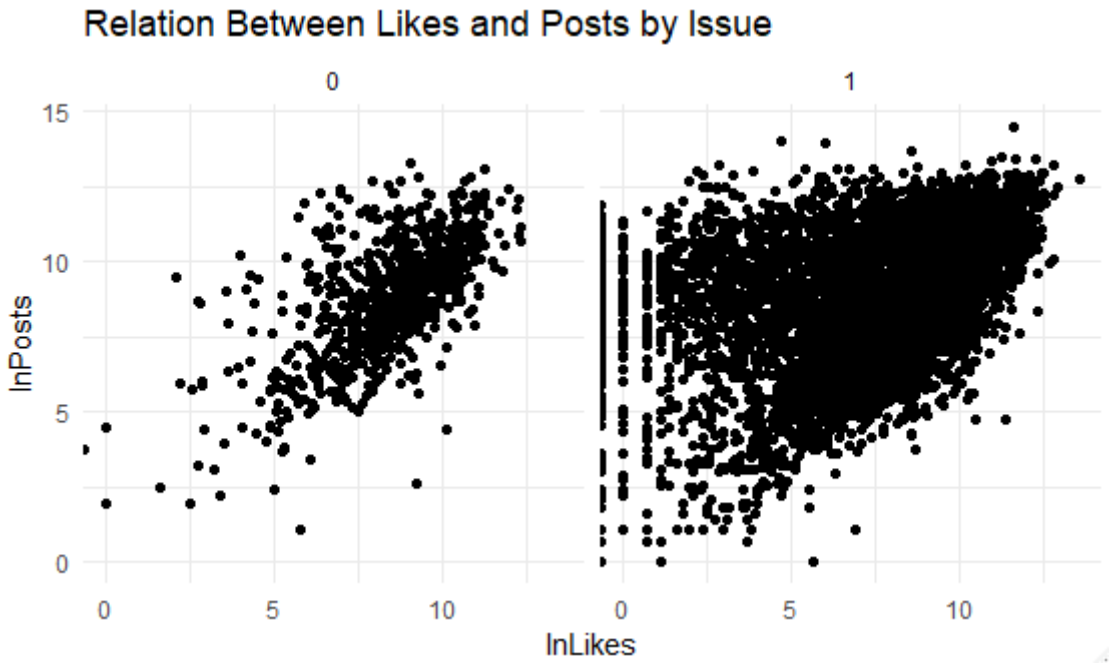

4000
3000
2000
1000
0
count
0
1
2
gender
evento
0
1

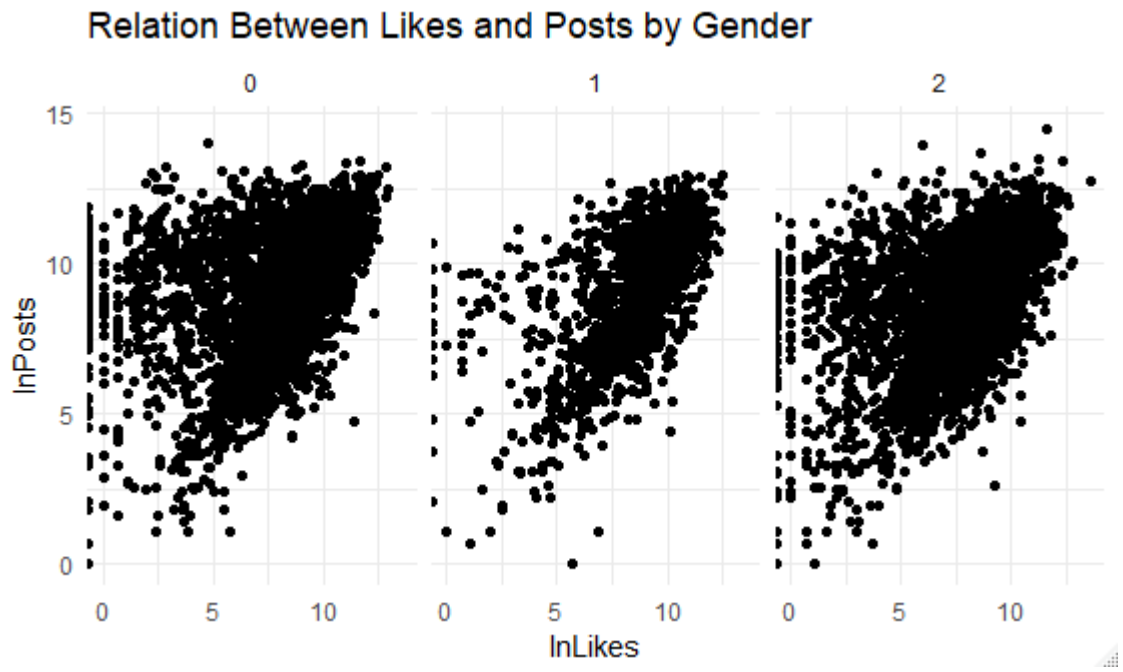

4000
3000
2000
1000
0
count
0
1
2
gender
religion
0
1

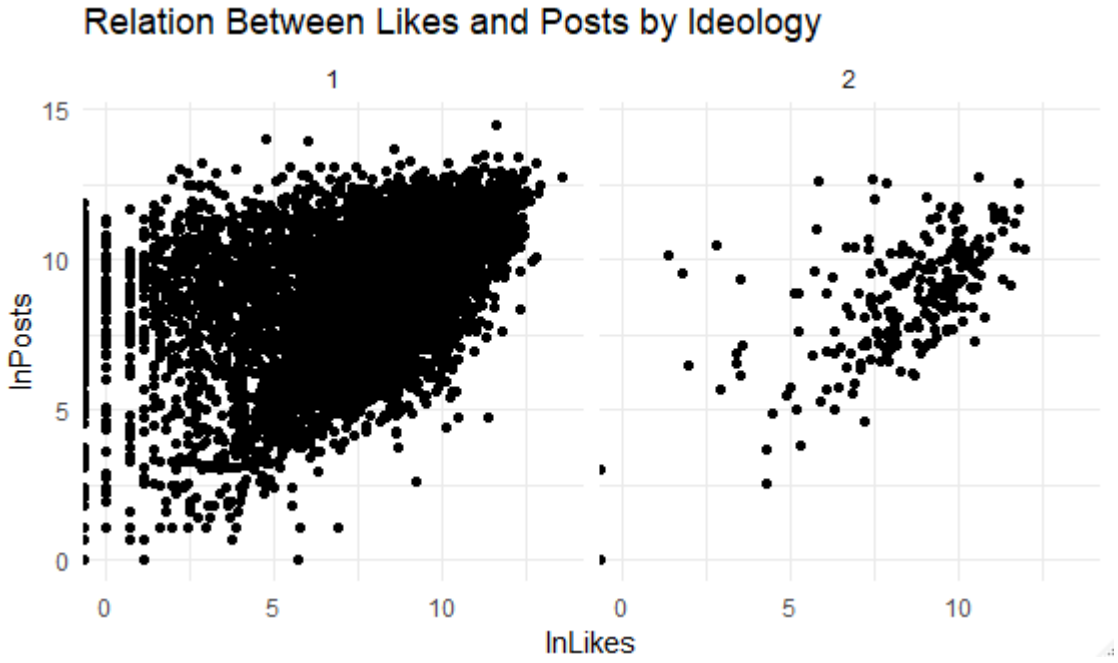


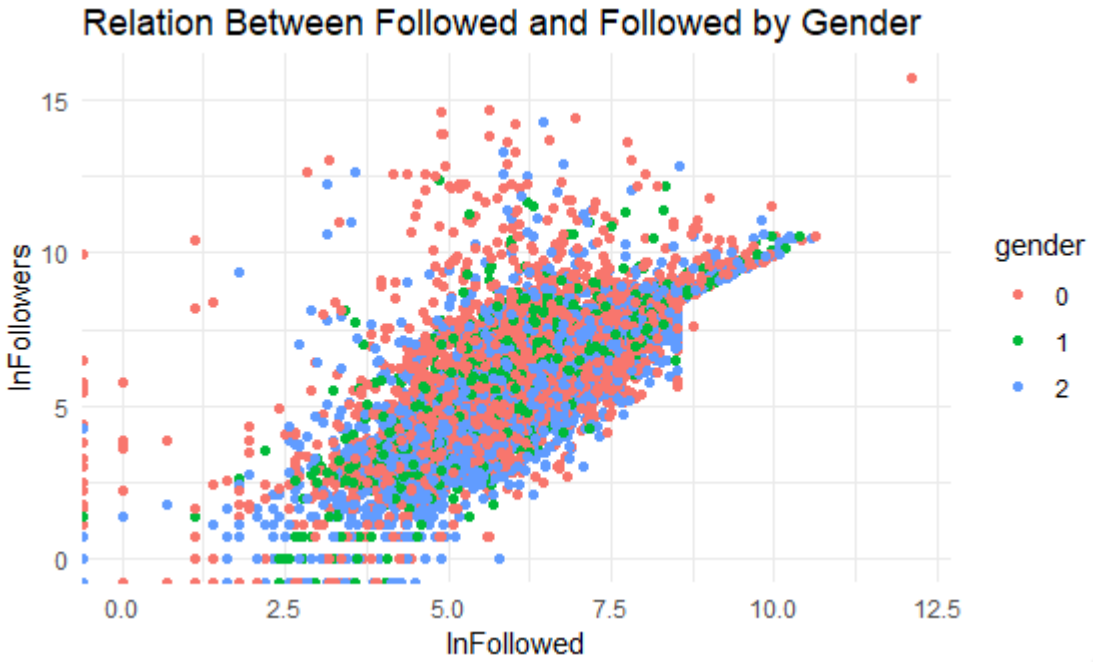
Relationship between the variables

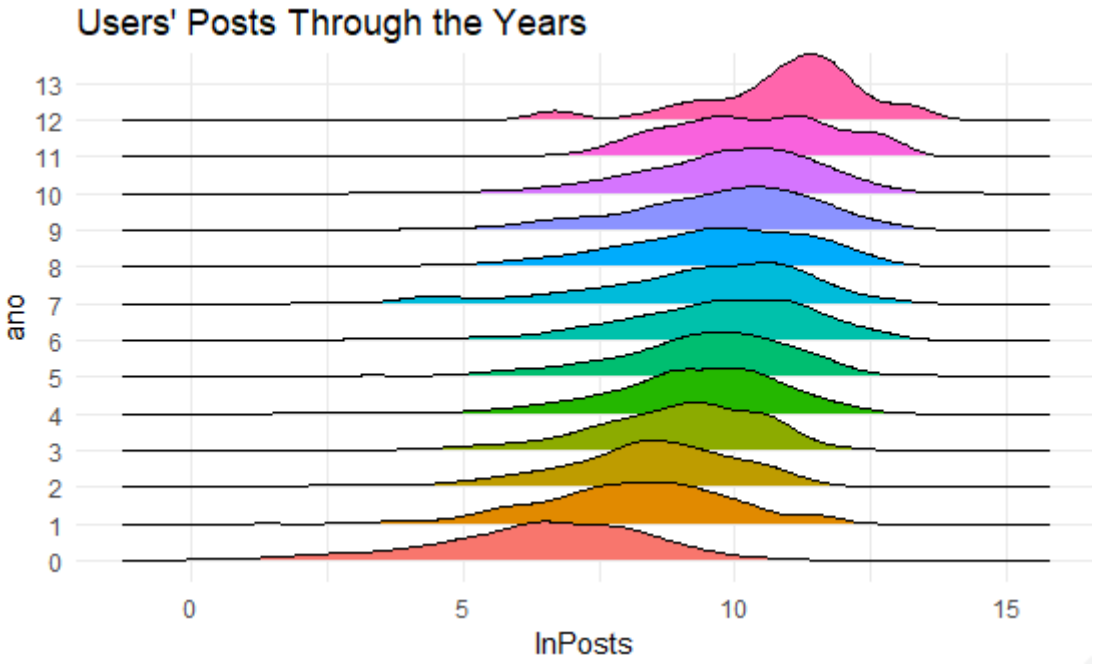

Relation Between Likes and Posts
lnPosts
lnLikes
ideology
1
2
0
5
10
15

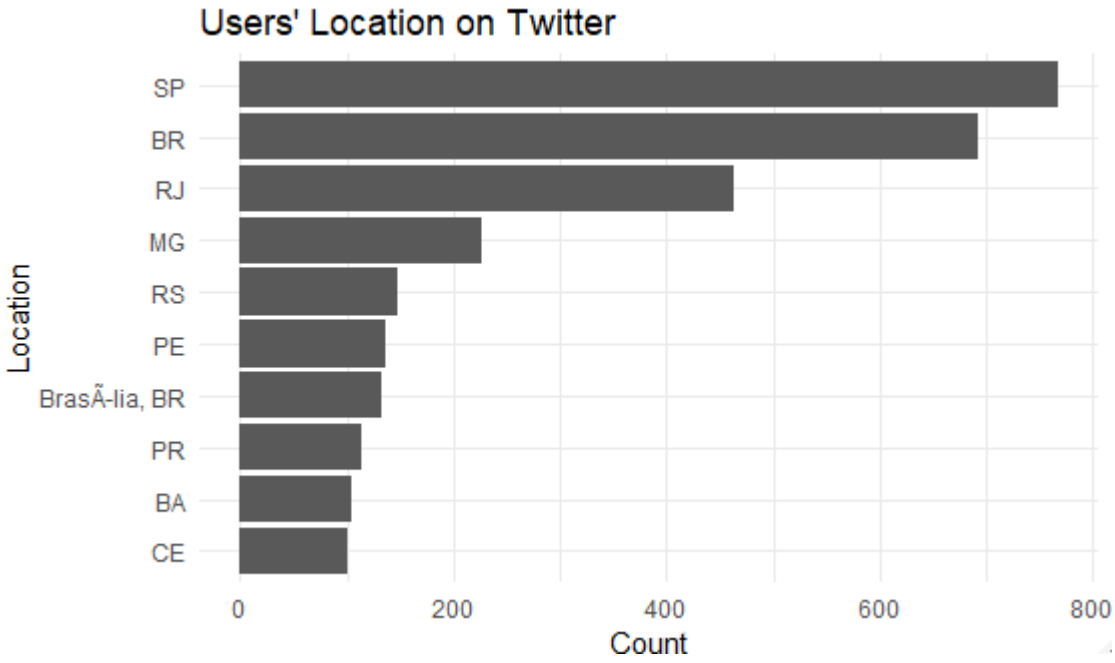

Relation Between Likes and Posts by Issue
0
1
lnPosts
lnLikes
0
5
10
15

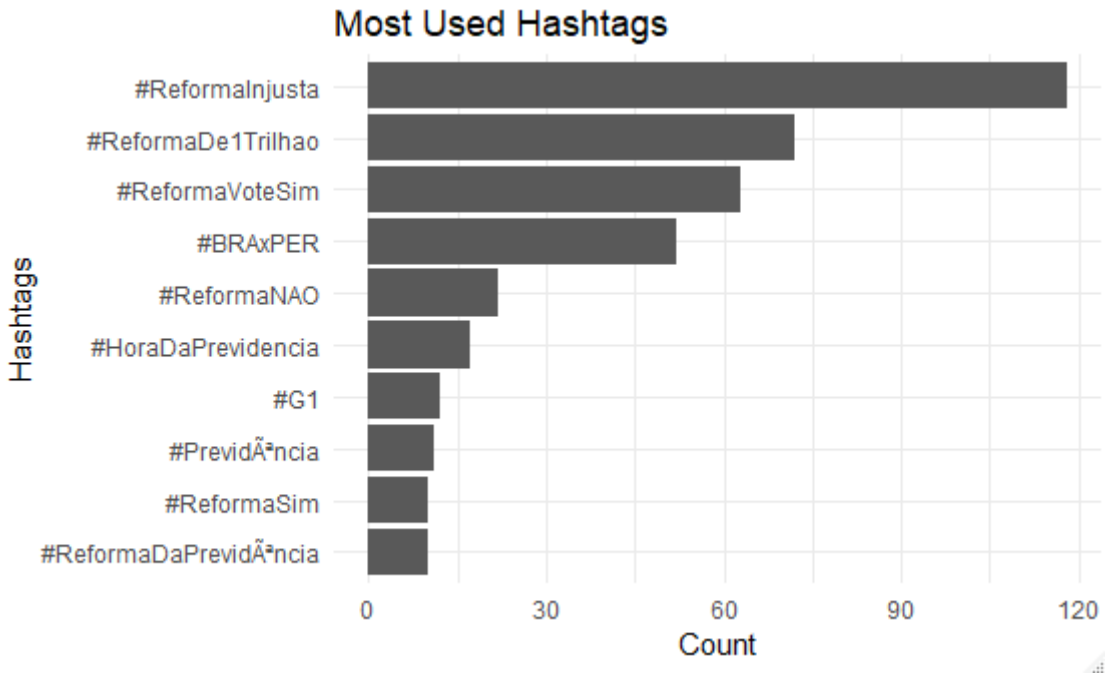

Relation Between Likes and Posts by Gender
0
1
2
15
10
5
0
lnPosts
0
5
10
lnLikes

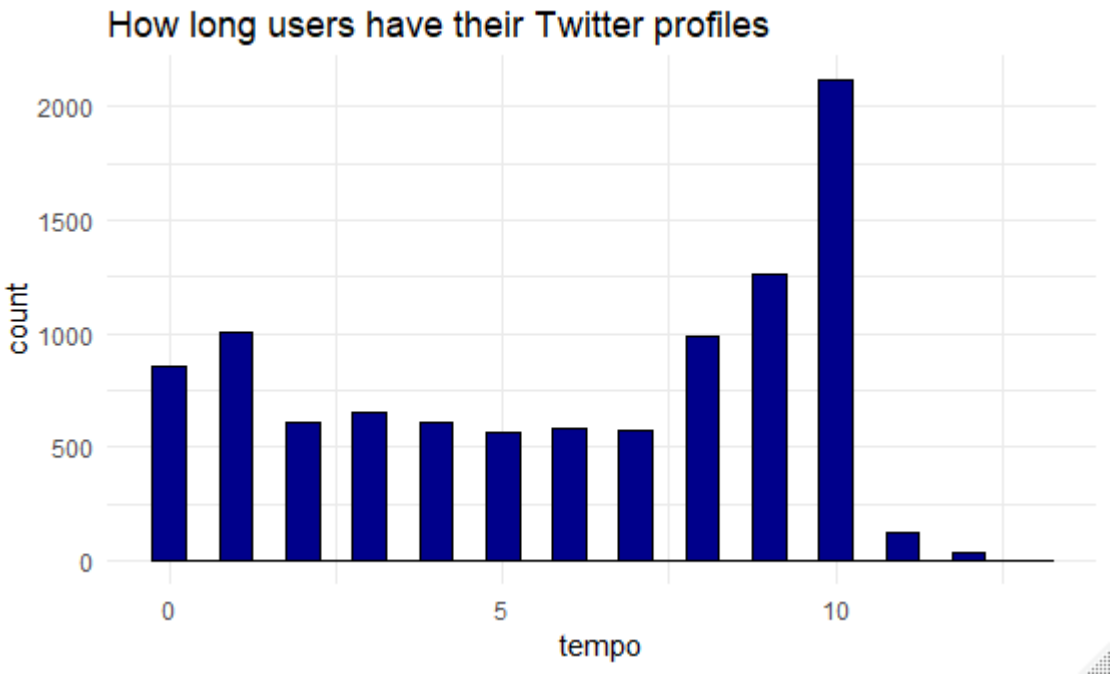

Relation Between Likes and Posts by Ideology
1
2
15
10
5
0
lnPosts
0
5
10
lnLikes


Relation Between Followed and Followed by Gender
lnFollowers
lnFollowed
gender
0
1
2
0
5
10
15
0.0
2.5
5.0
7.5
10.0
12.5


Users' Posts Through the Years
ano
lnPosts
0
1
2
3
4
5
6
7
8
9
10
11
12
13
0
5
10
15


Users' Location on Twitter
Location
SP
BR
RJ
MG
RS
PE
BrasÃ-lia, BR
PR
BA
CE
0
200
400
600
800
Count


Most Used Hashtags
Hashtags
#ReformaInjusta
#ReformaDe1Trilhao
#ReformaVoteSim
#BRAxPER
#ReformaNAO
#HoraDaPrevidencia
#G1
#PrevidÃªncia
#ReformaSim
#ReformaDaPrevidÃªncia
0
30
60
90
120
Count



Model Assumptions

We investigate collinearity by VIF (Variance Inflection Factor). It gives us na index measure of how much an estimated regression coefficient variance is influenced by collinearity. The higher this value, more collinearity issues to the model.

Model 1	**Vif**
N_followers	2.808676
N_followed	3.003366
N_posts	1.173434
N_likes	1.138439

Model 2	**Vif**
N_followers	2.818997
N_followed	3.013073
N_posts	1.233655
N_likes	1.138556
Issue_1	1.011086
Time	1.065111

Model 3	**Vif**
N_followers	2.823441
N_followed	3.024800
N_posts	1.174054
N_likes	1.138446
Ideology_2	1.007456

Model 4	Vif
N_followers	2.808676
N_followed	3.003367
N_posts	1.173602
N_likes	1.138691
Religion_1	1.000548

Model 5	Vif
N_followers	2.819450
N_followed	3.008035
N_posts	1.177203
N_likes	1.142459
Gender_1	1.133810
Gender_2	1.137275

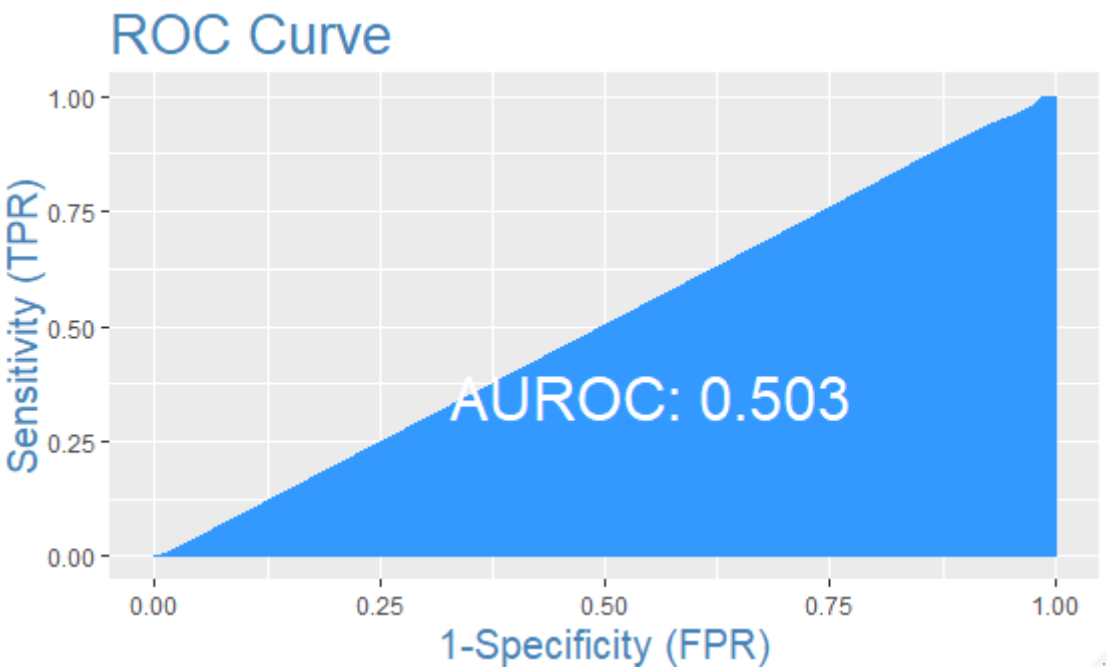
Model 6	Vif
N_followers	2.845773
N_followed	3.040072
N_posts	1.239329
N_likes	1.142878
Ideology_2	1.024359
Religion_1	1.014224
Gender_1	1.138999
Gender_2	1.154805
Issue_1	1.027151
Time	1.078658

Model Fitting

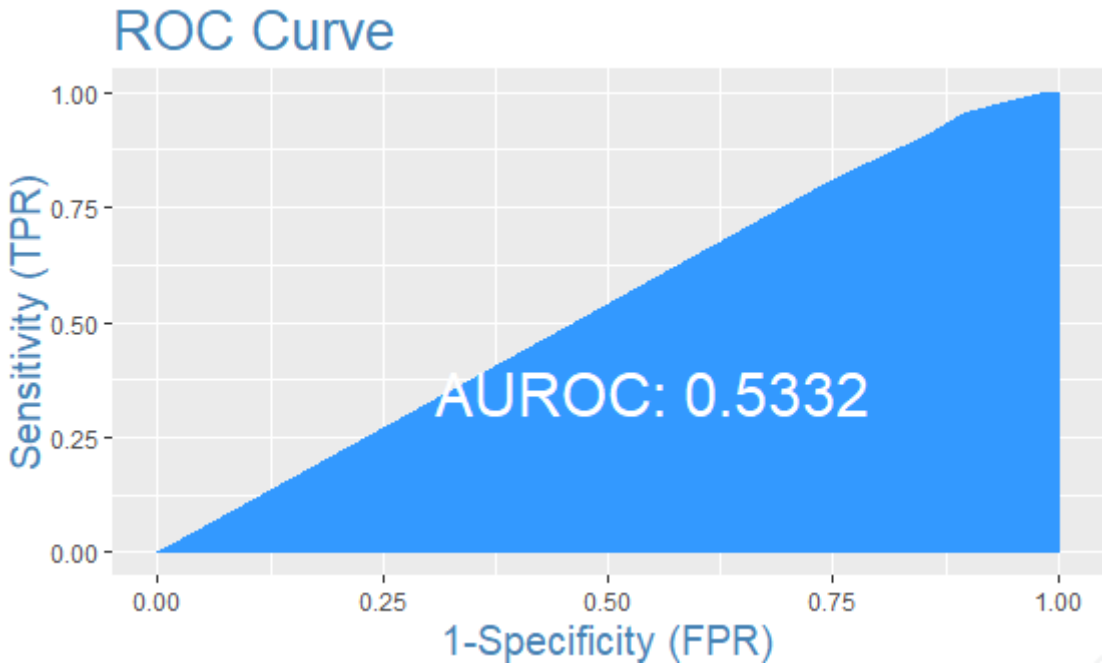
Logistic regression uses different fitting models. We will evaluate the models fitting though a statistical measure called Pseudo-R. This measure goes from 0 to 1. The closer to 1 means better explanatory capability.

Model	Pseudo-R^2
Model 1	0.00
Model 2	0.02
Model 3	0.00
Model 4	0.00
Model 5	0.01
Model 6	0.02
Model 7	0.02

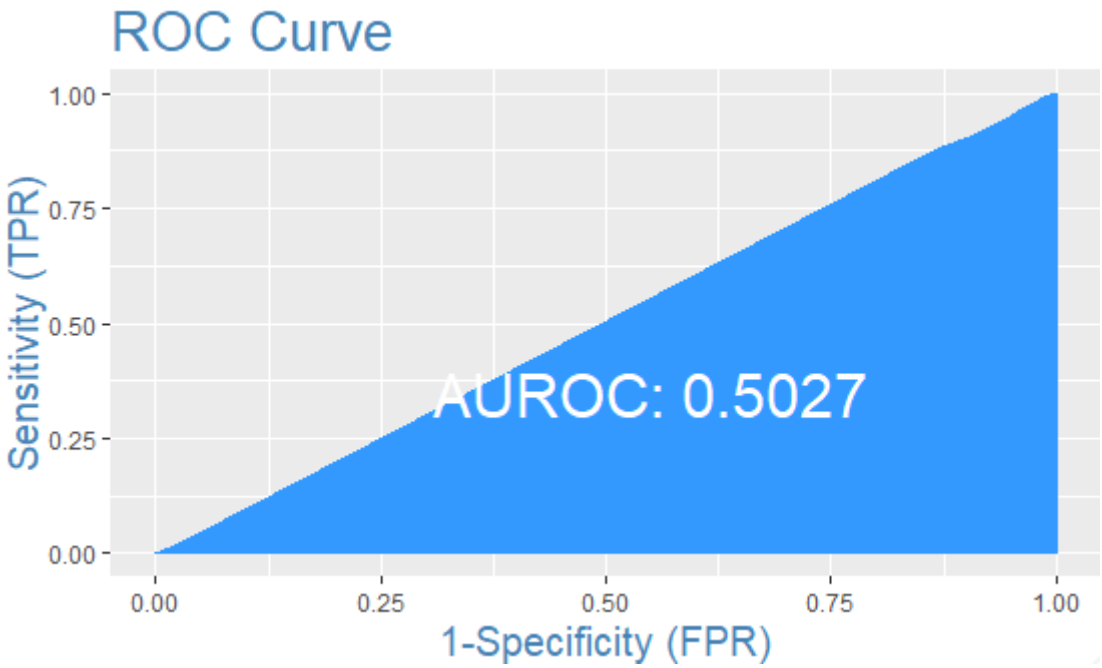
Model 1



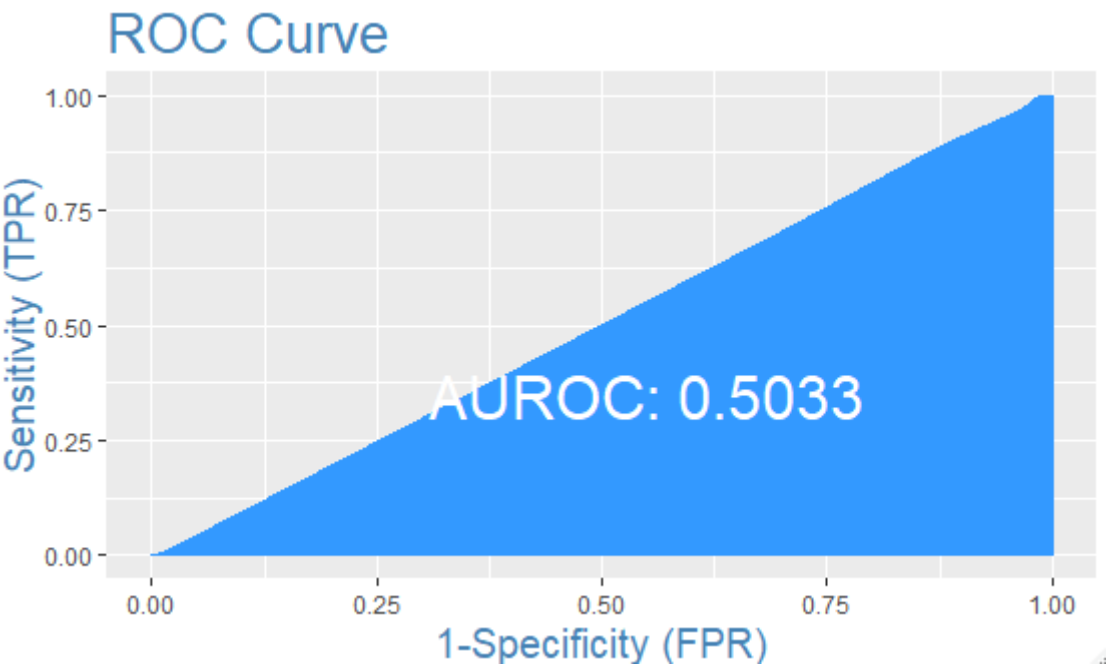
Model 2



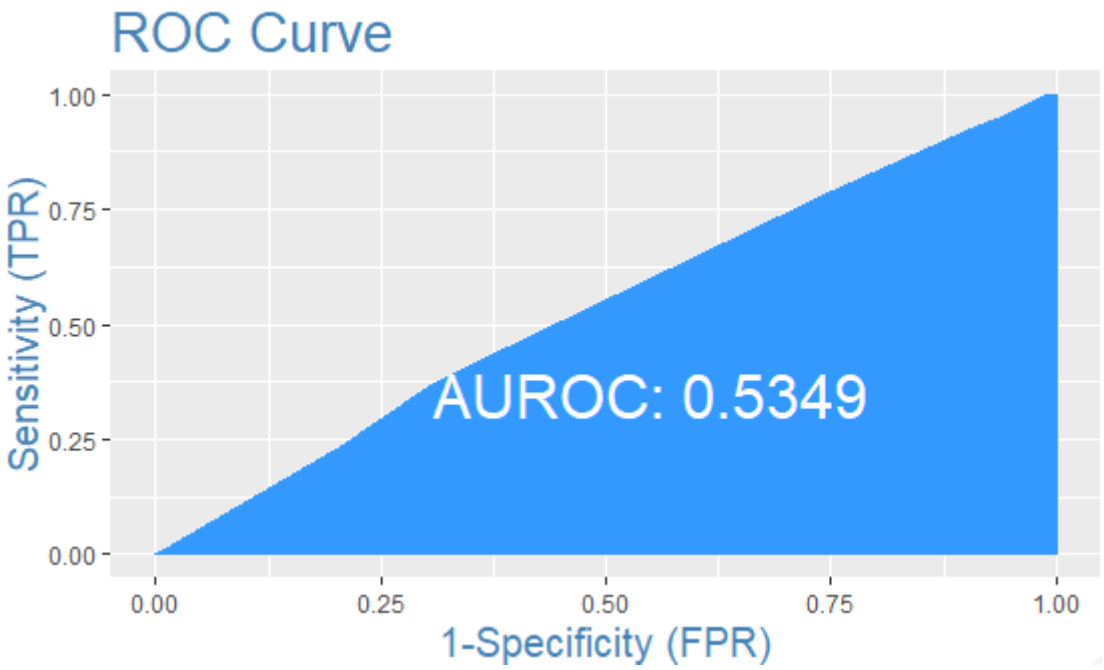
Model 3



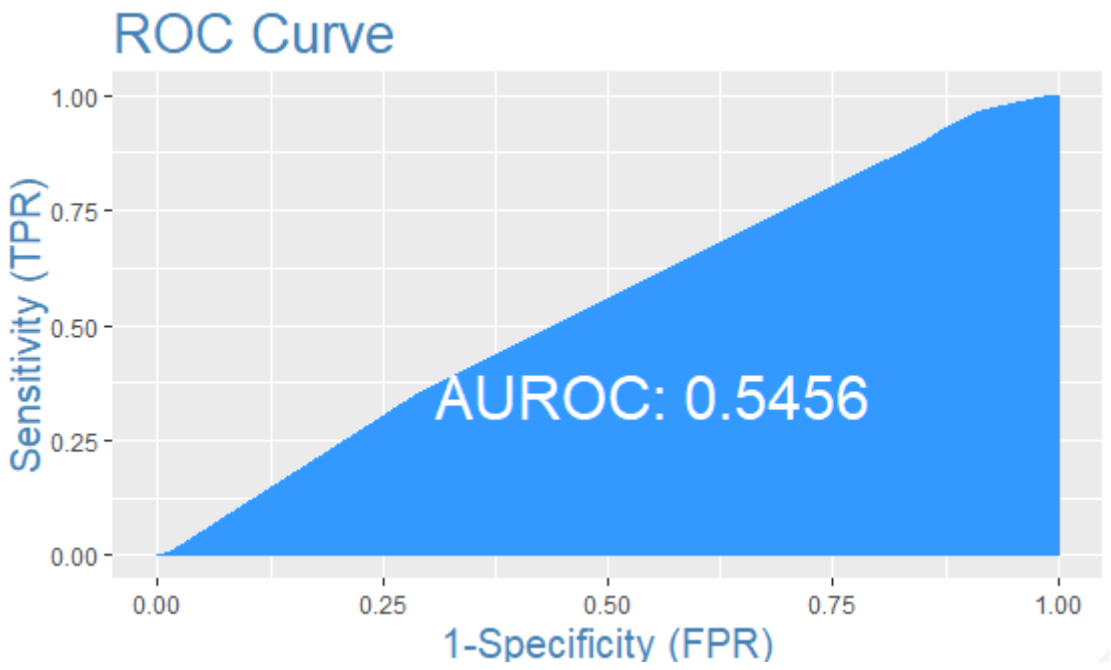
Model 4



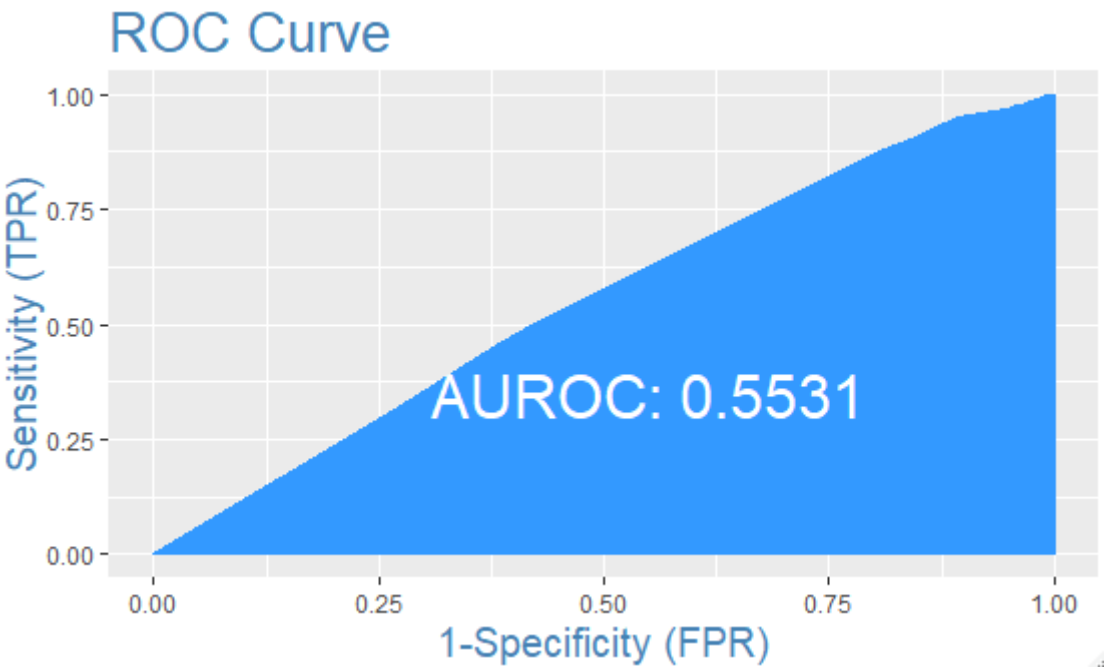
Model 5



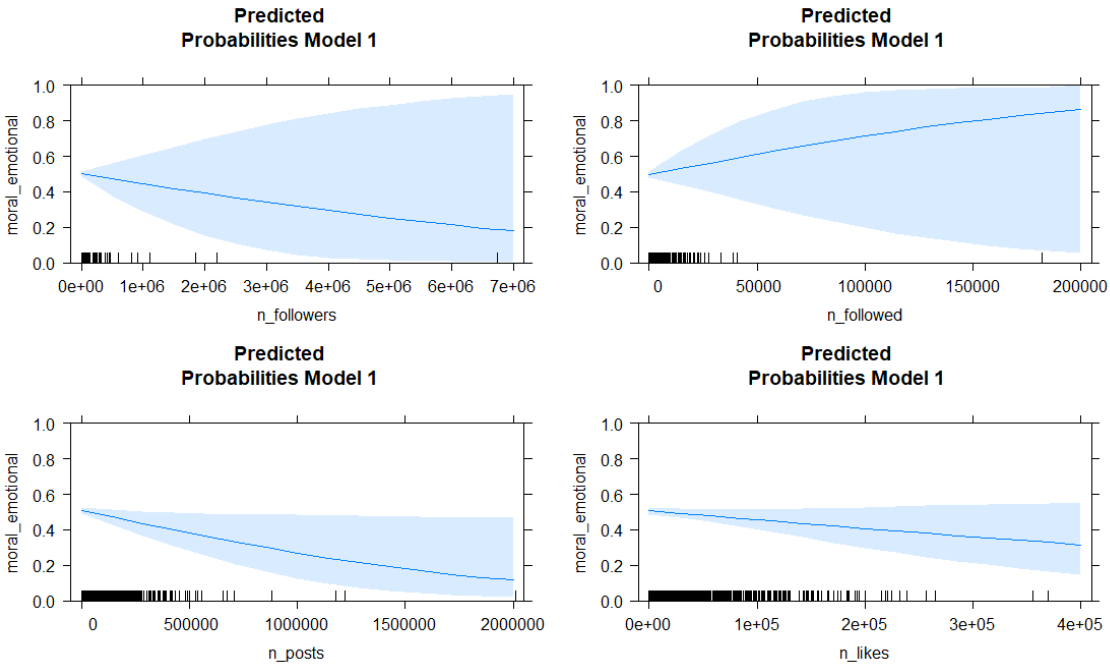
Model 6



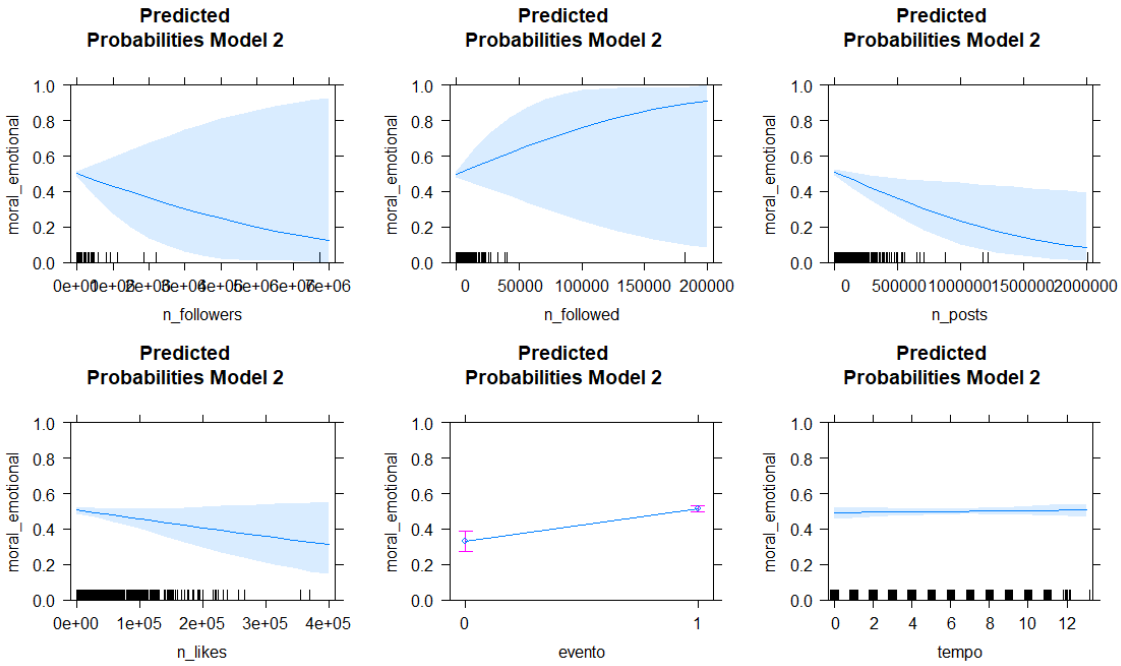
Model 7



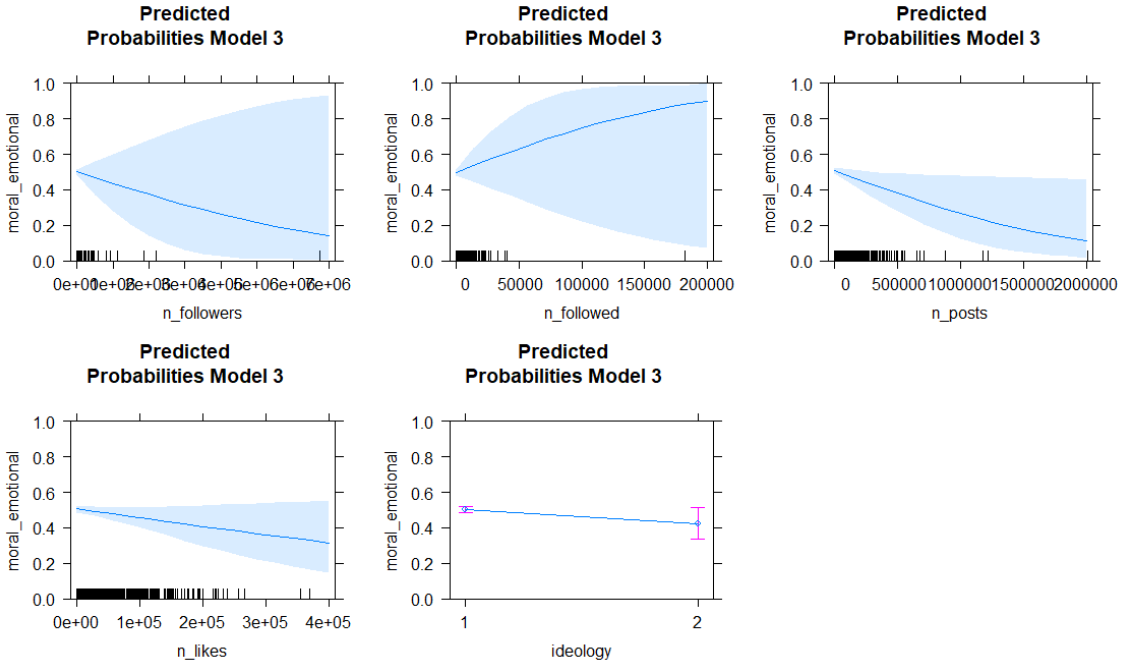
Predicted Probabilities: Model 1



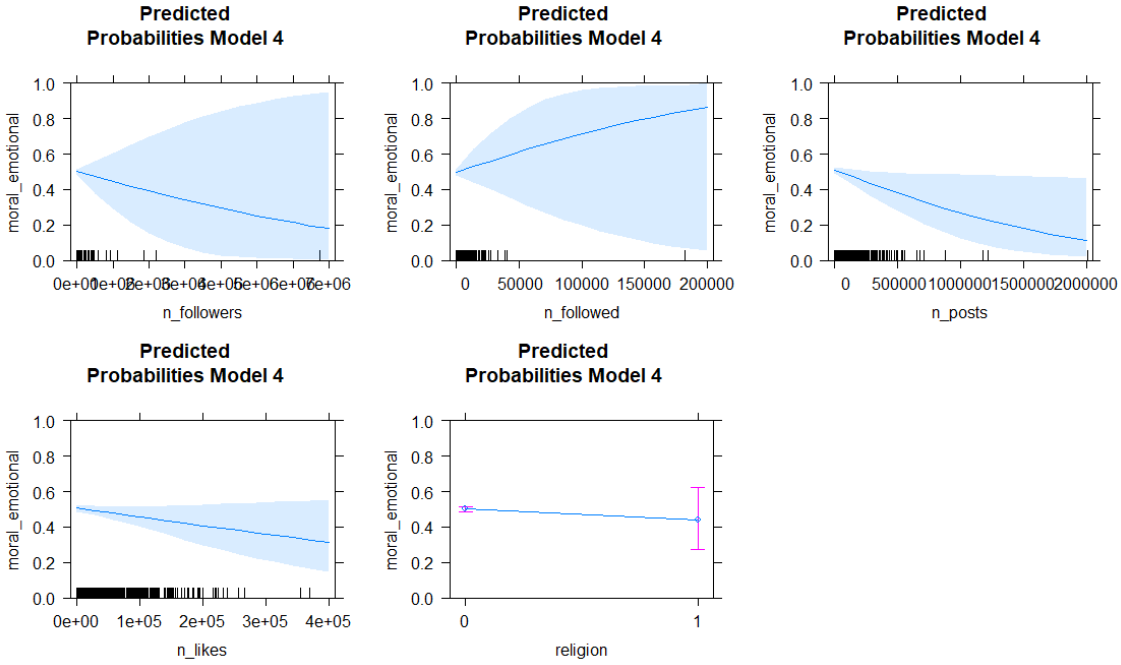
Predicted Probabilities: Model 2



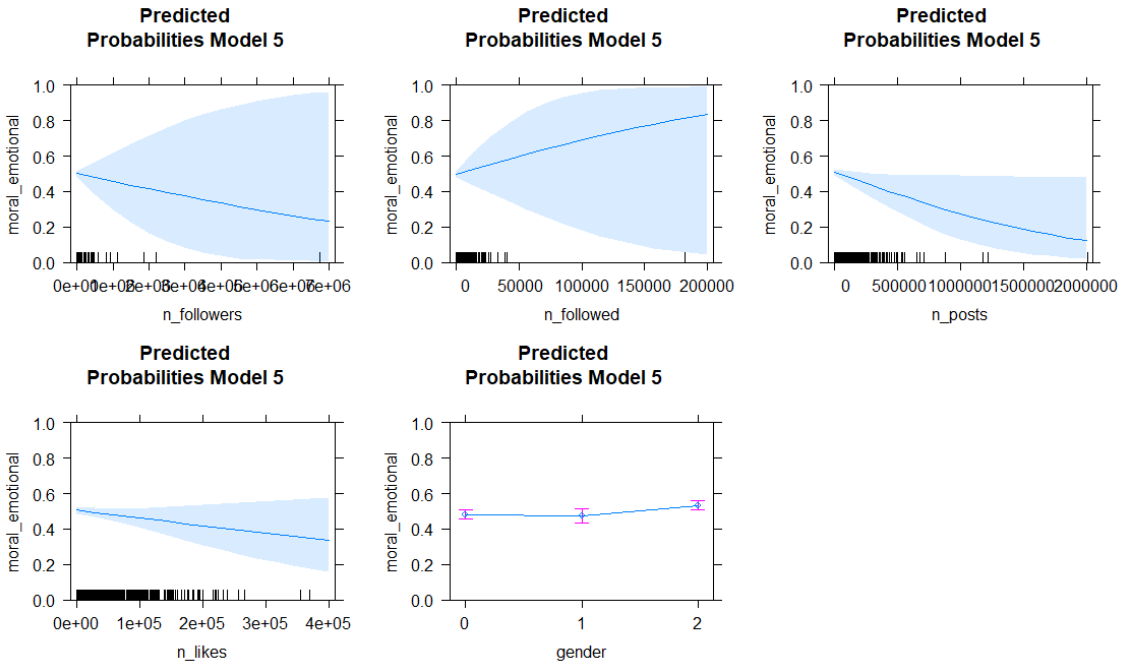
Predicted Probabilities: Model 3



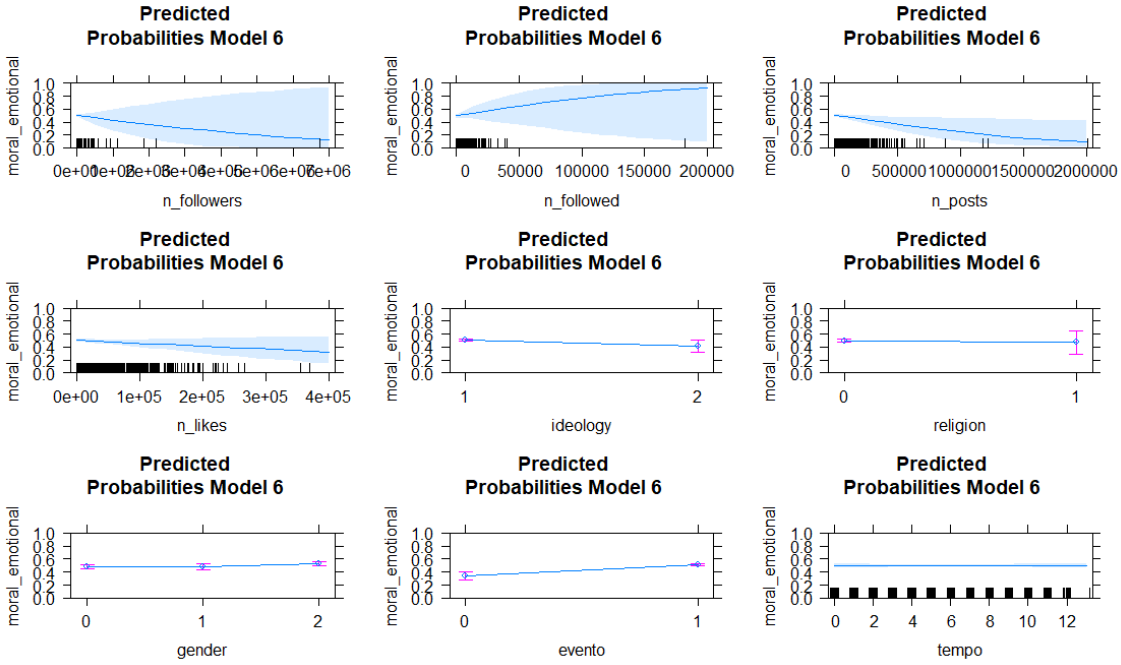
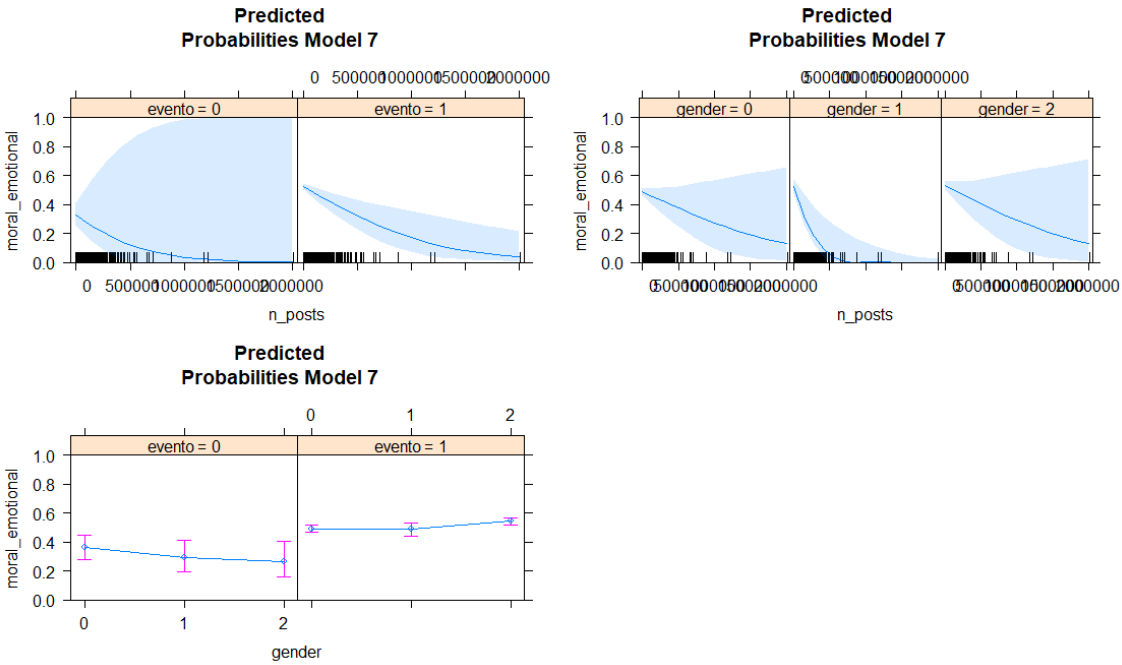
Predicted Probabilities: Model 4



Predicted Probabilities: Model 5



Predicted Probabilities: Model 6



Predicted Probabilities: Model 7