Liked in Facebook: The Correlation Between Facebook Popularity, Region And Populism

Nathália Viviani Bittencourt[1]

09/08/19

ABSTRACT

Does the Number of Followers on Facebook is Influenced by Populism and the Geographic Region of an Incumbent Leader? Given the growth of polarized elections and candidates who use social networks to deal directly with citizens, we seek to understand how these indices and the levels of populism in the world relate to each other. Especially as regards the elections of Narendra Modi in India, Donald Trump in the USA, and Bolsonaro in Brazil and Recep Tayyip Erdogan, there seems to be a similarity of conduct in looking at the values of their polarities in social networks and their categorization as populist government. The research by Team Populism and the English newspaper The Guardian, which identified populist levels in the world through the speeches of the rulers of each country, is used as a basis. (Not populist - somewhat populist - populist - very populist). Thus, the present work has as its main scope to understand how the popularity of leaders on facebook is higher according to their level of populism and geographical region. To this end, the methodology to be employed will be a multiple linear regression, by crossing data from the following variables: dependent, ie, the number of followers in social networks, notably facebook (face_likes), and the Independent Variables, "populist_very", "populist_somewhat populist, "America", Europe_Central Asia. As for the hypothesis to be tested, it is raised that popularity in social media has a strong relationship with governments considered populist. In addition, it is argues that the region of populist leaders is also strongly related to popularity on social networks (facebook). Finally, regarding the results, it couldn't be concluded that the high popularity in social networks is strongly related to the level of populismo, possibly because of the small number of observations we have in our data.

Palavras-chave: Populism, Facebook, America, Europe and Central Asia, Multiple Regression Analysis

[1] data access link for replication in GitHub: <https://github.com/Nvbittencourt/nathalia-bittencourt-ad-ufpe-2019>

[1] data access link for replication in Dataverse Project: < https://dataverse.harvard.edu/dataverseuser.xhtml?selectTab=dataRelatedToMe>

SUMMARY

1 INTRODUCTION

Some new populist tendencies arise under the Information Age. It is again a powerful political phenomenon, a scholarly topic, and even a complex word to describe a pattern in political discourse. Indeed, in the aftermath of the Great Recession that began in 2008, populism has returned to political life in a spectacular fashion and its has also become a trending topic in the global conversation.

From this perspective, some worldwide phenomena, such as the Brexit, as well as the Trump election in the USA, Recep Tayyip Edorgan in Turkey, Bolsonaro in Brazil and Narendra Modi in India, demonstrate similar attitudes towards communication, campaign strategies, and notably how they use social networks. Thus, the present work seeks to understand if and how this new context of populism endorses its characteristics through the digital media, especially through social networks.

It is important to note that transformations in the public sphere must be watched closely, specialy now that societies are increasingly medialized. Democratic communication is at the center of politics as never before, and there is a permanent war for winning over public perception in whichever issues are at the front of the public agenda. While the mass media adhere to professional norms and news values, social media serve as direct linkage to the people and allow the populists to circumvent the journalistic gatekeepers. In this way, social media provide the populists with the freedom to uncontestedly articulate their ideology and spread their polarized messages of the Good *vs* Evil, Elites *vs* The citizens and a stress to the sovereignity of the people.

Besides, even if some theoretical relation between populism and online communication was already established early in the history of the Internet (Bimber, 1998), there are just a few works that refer specifically to social networks as a form of persuasion in politics and as a way to perpetrate ideologies. In this sense, some scholars bestowed the Internet with the potential to restructure political power in a populist direction and to promote unmediated communication between politicians and citizens

However, a recent article written by Sven Engesser, Nicole Ernst, Frank Esser & Florin Büchel (2017), for instance, investigates whether and how populism manifests itself on social media They conducted a qualitative text analysis of typical Facebook and Twitter posts and took four West European countries into account: Austria, Italy, Switzerland, and the UK in order to increase the variance and to identify cross-national patterns. Thus, it is essential to increase the scope of studies on the relationship between the use of social media and the populist rise in the world. In the scope of this work,

In this perspective, this paper seeks to analyze the relationship between various incumbent leaders and their respective number of likes on facebook. Is it possible to assert that the more popular an incumbent leader is on social media, the more populist he or she is? I also raise the hypothesis of the region as a variable that influences the correlation result

To sum up, I believe that the present paper will increase the discussion about the intense use of social media as a common source of communication to populist leaders. It is intended to analyze if the high number of followers is influenced by the use of populist leaders and the region where it is located.

2 LITERATURE REVIEW

One of the first attempts to conceive of populism as a uniform phenomenon dates back to Shils (1956). Although populism has been part of the academic debate since then, it is still a contested concept and has been described as a ‘notoriously vague term’ (Canovan, 1999).

Halkins and Castanho Silva (2016) make a theoretical review of the studies of scholars on popuism in the world, and find that most do not seek cross-sectional analysis, but rather of small countries that have common characteristics in their discourses. Thus,in their work they try to bring a more broadly definition to the literature by indicating what a populista discourse should be like:

> It is a discourse which sees politics as divided in moral terms, where the good is identified with “the people” and the evil is embodied by an “elite”. This “people” encompasses the majority of the population and is a homogeneous, unified body that has an identifiable will – the General Will or volont´e g´en´erale –, which should be guiding all decision-making in politics The elite, on the other side, is a minority who is in power (or in risk of imminent return to), who uses its resources to exploit the people. It is morally evil, and to blame for all bad things that befall the country.[2]

The above work has been important in understanding how populist discourse can follow similar patterns around the world, regardless of region. However, at least three other papers were fundamental for the present research idea to be raised.

Regarding the relationship of the intense use of social media by populist governments, there are few works that raise the influence of twitter and facebook as easy platforms for the raise of populist discourse. In this sense, the work cited by Sven Engesser, Nicole Ernst, Frank Esser & Florin Büchel (2017) is groundbreaking in moving away from the commonplace of traditional media and studying the new forms of communication of the 21st century. Indeed, this work encouraged me to test the relationship between the high number of followers on Facebook and its relationship with populist governments.

Moreover, Fiona's work (2019), which studies the volume of populist governments and its geographical role has served to raise my second hypothesis, namely, whether the number of followers on twitter can be influenced by the leader's region The scholar illustrates below some patterns that influence the increase of populism in different regions of the planet:

---

[2]HAWKINS, Kirk A., and SILVA, Bruno Castanho, 2018. **"Text Analysis: Big Data Approaches"**. In: The Ideational Approach to Populism: Theory, Method & Analysis, edited by Kirk A. Hawkins, Ryan Carlin, Levente Littvay, and Cristóbal Rovira Kaltwasser. London: Routledge.

> The first argument is centred on issues of economic inequality. According to this thesis, populista voters are the 'victims of neoliberal globalisation',21 with its attendant impacts on international competitiveness, sectoral specialisation, and income and job security, together with the uneven effects of the fallout from the financial and economic crisis on interpersonal disparities. The second argument is that the key drivers are not material but cultural or attitudinal, linked to 'disenchantment with the broken promises of liberal modernity', 'new forms of social marginalisation and the notion that some 'progressive' values – such as multiculturalism, gender, race and LGBT equality - have simply evolved too far and too fast for comfort. Importantly, for this discussion, patterns of populism are not spatially neutral and accordingly the geographical dimension to populism and its causes is also receiving growing attention[3]

In short, MOFFIT's (2016) book about the global rise of populismo addresses the growth of populist leaders in some regions of the planet, and reiterates a specific breakthrough in the Americas and Europe. In this sense, this publication was an incentive for the present work to also take into consideration the region of the country's incumbent leader to our hypothesis, which will be better explained in the next section.

[3] WISHLADE, Fiona, 2019. **The Rise of Populism, Regional Disparities and the Regional Policy Response.** European Policy Research Paper. University of Strathclyde Publishing. Available at: <https://www.google.com/search?q=the+rise+of+populism%2C+regoional+disparities&oq=the+rise+of+populism%2C+regoional+disparities&aqs=chrome..69i57j69i64.12227j0j4&sourceid=chrome&ie=UTF-8>. Accessed in 8 of August, 2019.

3 THEORY

This paper seeks to understand if there is any relationship between the number of followers on social networks, notably facebook, and the fact that the country has a populist leader. Is it possible to state that the more followers in the incumbent governant's facebook account, the more likely he has a populist discourse?

In addition, we try to test the following assumption: is there any influence on the region of the populist leaders that explains the high following? So these are the two assumptions I should test. To come up with these ideas, the literature from the previous section helped to formulate that **a leader's facebook following increases with his level of populism and according to the geographical region.**

To reach this conclusion, some world phenomena were observed. Indeed, the intensive use of social networks as a campaign strategy in the last US election, the rise of Bolsonaro in Brazil and its similarity to Trump's *modus operandi* during brazilian elections, the Turkish President, Recep Tayyip Erdogan, his controvertial re-election and intensive use in his personal account, Narendra Modi's massive use of social media during his recent campaign, and so on.

Finally, the recent research by The Guardian newspaper in partnership with the Team Populism group served as a strong support for this work. Indeed, their research is compiled in the Global Populism Database, the most comprehensive and reliable tracker of populist discourse in the world. They attributed an average populism "score", based on the extent to which speeches contained populist ideas. The average populism score, across all 40 countries, has doubled from 0.2 in the early 2000s to around 0.4 today. The number of countries with leaders classified as at least "somewhat" populist – a score of 0.5 and above – has also doubled in that period, from seven in 2004, to around 14 in recent years.

Thus, the growth of populism in the world, as the Guardian survey in partnership with Team Populism concludes, has served as na assumption to support that the presente work is in line with a global trend. In the next section, I fully explain how I built my model with the data from the recent research by The Guardian with the Team Populism.

4 RESEARCH DESIGN AND DATA

During the research design formulation process, an observational cross-sectional study was carried out, as the objective is to analyze how the variables behave from the regression method. In this regard, I searched the database of the above research which measures the level of populism among various leaders around the world. To illustrate, The Guardian newspaper explained the methodology involved in creating the populist score.

> Hawkins and his colleagues created a system that apportions a “score” for speeches based on the extent to which they contain populist ideas, discourse and rhetoric. The scores run from 0 (no populism – “**not populist**” speech category) to 1 (clear populism, but used inconsistently or with a mild tone, which is called **“somewhat populist**” speech category) to 2 (clear populism used consistently with a strong tone – “**populist**” or “**very populist**” speech category). “Clear populism” means that the core elements of populism – the notion of a virtuous will of the common people and the notion of an evil, conspiratorial elite – are present..[4]

It’s importat to notice that recent elections in Mexico and Brazil also posed problems for researchers to analyse their populismo score. Neither Andrés Manuel López Obrador nor Jair Bolsonaro have been in office long enough to provide researchers with a wide sample of non-campaign speeches. The sample that was coded may not accurately reflect the rest of their time in office.

As mentioned, populism levels were classified as “very populist”, “populist”, “somewhat populist” and “not populist”. During the treatment of these variables in R program, we made two transformations so that they were adjusted to what we want to analyze. I created an independent variable with only the “populist” and “very populist” scores, which we call "Pop_VeryPop".

In addition to that, we created another Independent variable named "Pop_VeryPop_SomewPop", which is the first one added to the "somewhat populist" score. I decided to create this to analyze its influence on the results, as this kind of score has duplicated. Durng the research’s time, according to the results from The Guardian and Team Populism. I also used the independent variable Not_Populist from the original data. I made 3 linear regression models with these independent variables in order to analyze the direct

[4] LEWIS, Paul, CLARKE, Seán and BARR, Caelainn. ‘**How we combed leaders’ speeches to gauge populist rise.** The Guardian Newspaper. London, 6 March 2019.Avalable at <https://www.theguardian.com/world/2019/mar/06/how-we-combed-leaders-speeches-to-gauge-populist-rise> Accessed in july, 2019

correlation with the dependent variable (face_likes) before moving on to the main mutivariate models.

Finally, we use two variables that will endorse the main purpose of this research, which is to analyze if the number of followers on facebook is influenced by populism and the geographic region of the incumbent. Thus, two variables were created: "America", which takes into account the leaders of the entire American continent (North America, the Caribbean and South America) and the variable "EuropeCentralAsia", which deals with the leaders from that region. The Pop speech Data from the Guardian and Team Populism research.[5]

Let's move on to explaining the creation of our dependent variable, "face_likes". It represents the number of followers of each incumbent personal account. To do so, we accessed the data "World Leaders on Facebook 2019[6]", available on the website of Twiplomacy Organization for download. In addition, in order to make data processing more secure, we accessed the data "World Leaders on Social Media"[7], which presents the leader popularity on various social networks, but I only managed the facebook data.

Given this, we created a dependent variable called "face_likes" and 5 independent variables, but only three of these were used to fit the Multivariate Regression model, which are: "**Pop_VeryPop_SomewPop**", since we want to measure if populism is able to influence the number of Facebook followers, "**America**" and "**EuropeCentralAsia**", which refer to the leader's geographic region. Thus, we made two multivariate regressions: **RegM1_America**, which analyzes facebook followers from the influence (if any) of the "**Pop_VeryPop_SomewPop"** variable and the "**America**" region.

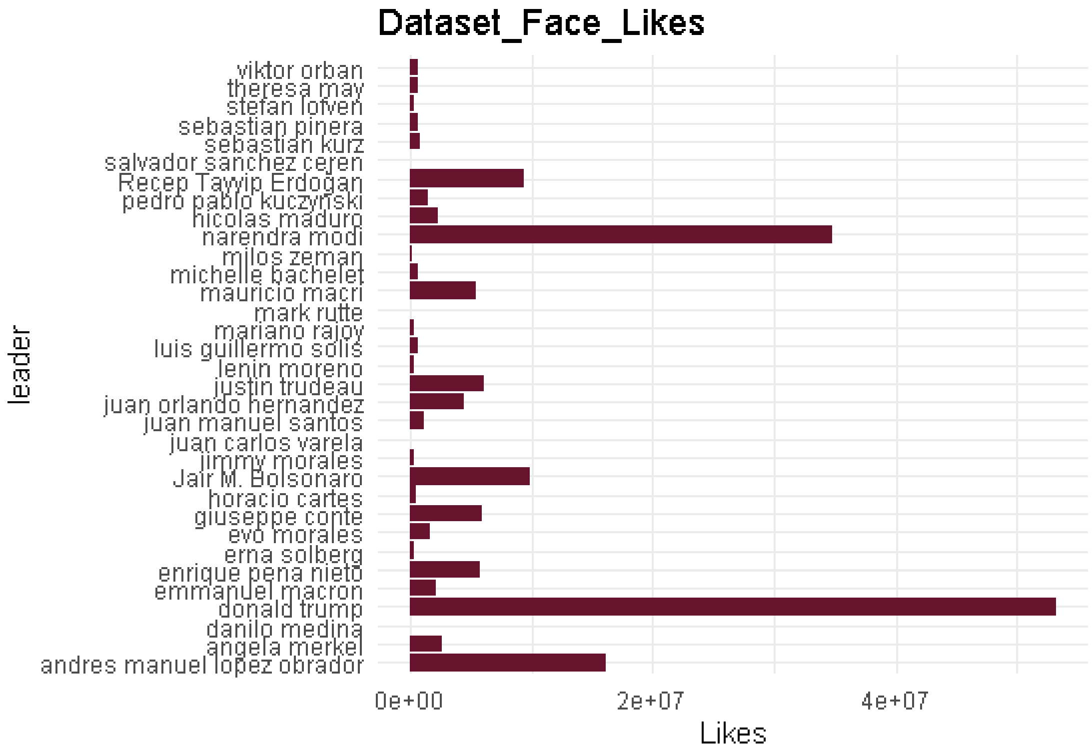
The second, "**RegM2_Europe_CentralAsia**", is the multivariate regression that attempts to explain "**face_likes**" through the independent variables "**Pop_VeryPop_SomewPop**" and **EuropeCentralAsia.**

The whole treatment code is available in "Methodological Appendix" section. From the data treatment we managed to creat our Dataset, that has 33 observations and 5 variables. In the next section, I analyse the models with tables and graphics.

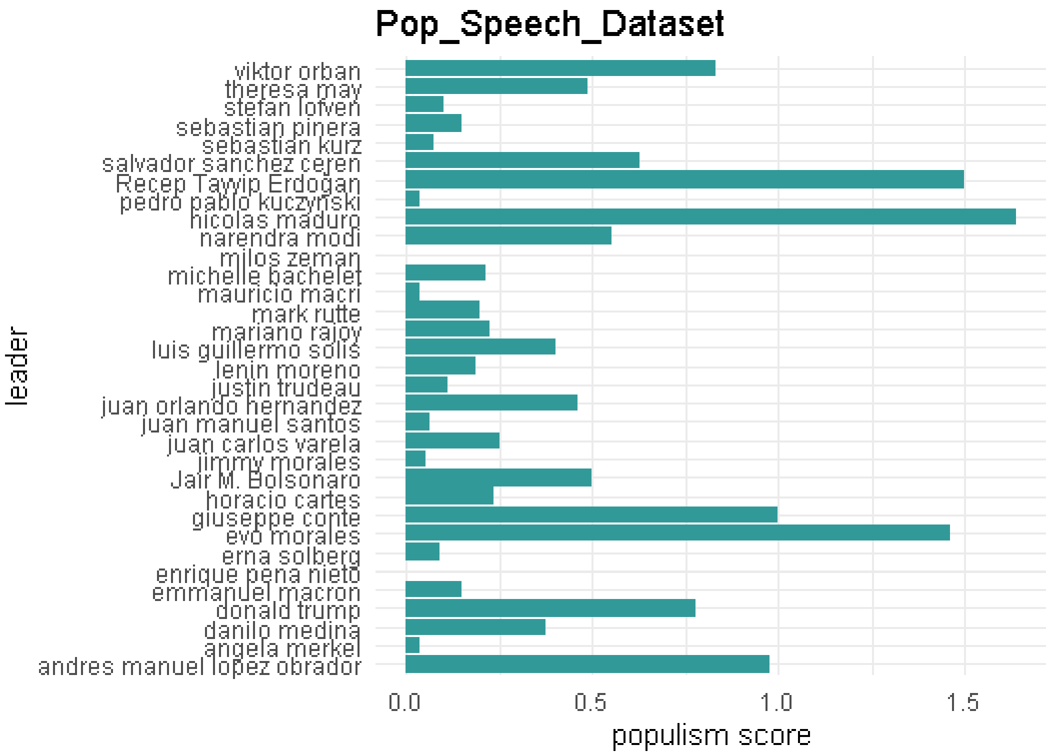
[5] Populist speech Data.xlsx is available to download at <https://populism.byu.edu/Pages/Data>
[6]Twiplomacy-Facebook-2019-Data-File.xlsx is available to download at <https://twiplomacy.com/blog/world-leaders-facebook-2019/>
[7] World_leaders_followers.xlsx is available to download at <https://www.cuponation.com/world-leaders-social-media/>

5 RESULTS

Now that we have generated our dataset, we can now show some graphs that allow a better view of the table. From the next column graph, we're able to see how many likes the leaders generated from the data treatment have.



Right away, we spot two Presidents who stand out on facebook social networks: Donald Trump and Narendra Modi. According to the next column graph (pop speech score) we can see that they're considered to be scored as somewhat populists. Jair Bolsonaro (South America), Recep Tayyip Erdogan (Europe and Central Asia) and Andrés Obrador (Central America) also feature prominently on the chart



As we can see on the column chart above, a lot o leaders feature as somewhat populist (0.5) or a higher score. This may mean that there is a tendency for populism to increase in some regions.

Furthermore, as mentioned in the previous session, we made three linear (univariate) regressions so that we could analyze whether there is a direct correlation with the dependent variable face_likes. Therefore, the table below illustrates the following results:

Results

	face_likes		
	Model 1	**Model 2**	**Model 3**
|Populist + Very Pop.|	2535321.000		
	(5893276.000)		
|Populist + Very Pop. + Somew. Pop.|		10156032.000***	
		(3571462.000)	
|Not Populist			-10156032.000***
			(3571462.000)
Constant	4793461.000**	1407670.000	11563702.000***
	(2051775.000)	(2153672.000)	(2849041.000)
N	33	33	33
Log Likelihood	-581.983	-578.257	-578.257
AIC	1167.966	1160.513	1160.513

***p < .01; **p < .05; *p < .1

As we can see in this chart, only the second and third models have any level of significance so that p values are less than 0,01. The 'AIC' parameter means "Akaike information criterion", which is an estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models.

Because our sample size is small (dataset has only 33 obs.), there is a substantial probability that AIC will select models that have too many parameters.As the AIC parametercompares the models available, then model with lower AIC would be the preferred one. So we can affirm the second and third modelhave a more impacto n the dependente variable. Given that the number of cases is small, we cannot safely say that the second and third suggested linear regression models are significant. We can only state, according to the table, that the first model is less satisfactory than the others. For more observations, some graphics are available in the methodological appendix.

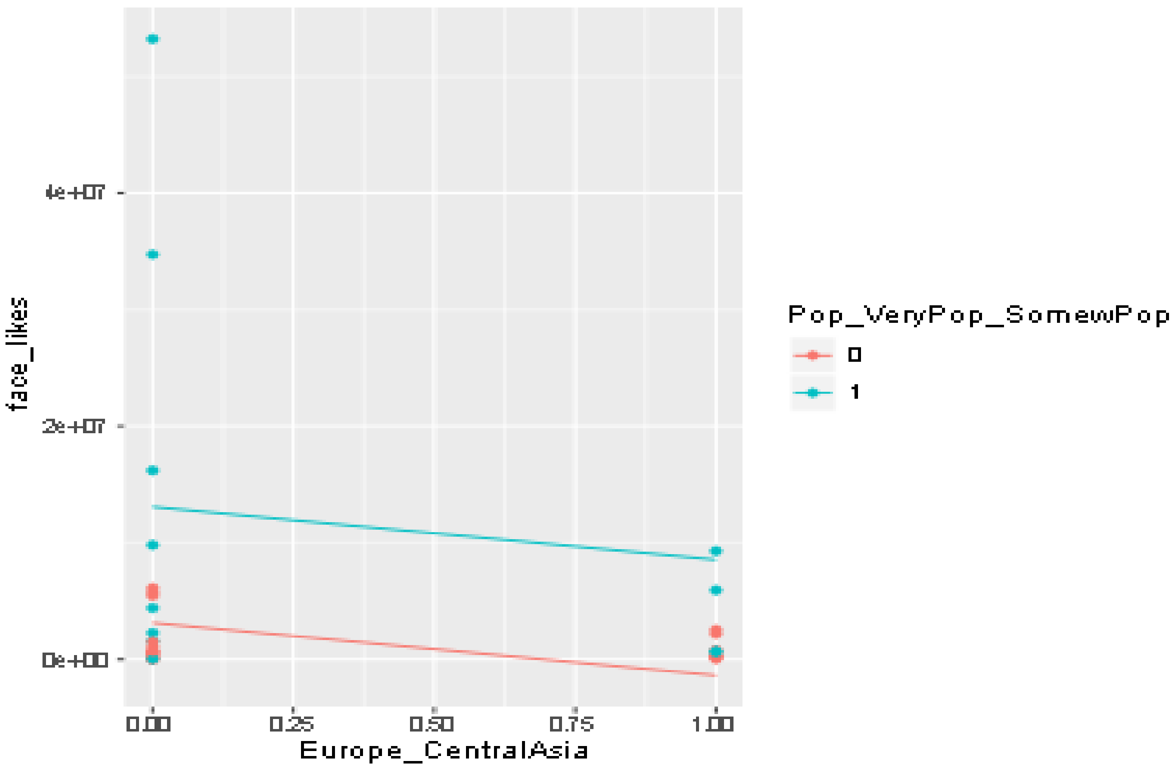
The next table analyses the two mulivariable regressions we proposed in our work. As we can see below, R-squared values explain around 21% and 24% of our models, respectively. It's importante to understand that R-squared percentage evaluates the scatter of the data points around the fitted regression line. It is also called the coefficient of determination, or the coefficient of multiple determination for multiple regression. High R-squared values represent smaller differences between the observed data and the fitted values. Again, our populist variable (Pop_VeryPop_SomewPop) presents p-value under 0,01, which means that it is signifcant to the dependente variable. However, it doesn't seem to exist, as our data (Dataset) suggests, a correlation between athe variables altogether in our model. Again, we emphasize that this may be happening because our data doesn't have many observations.

Results 2

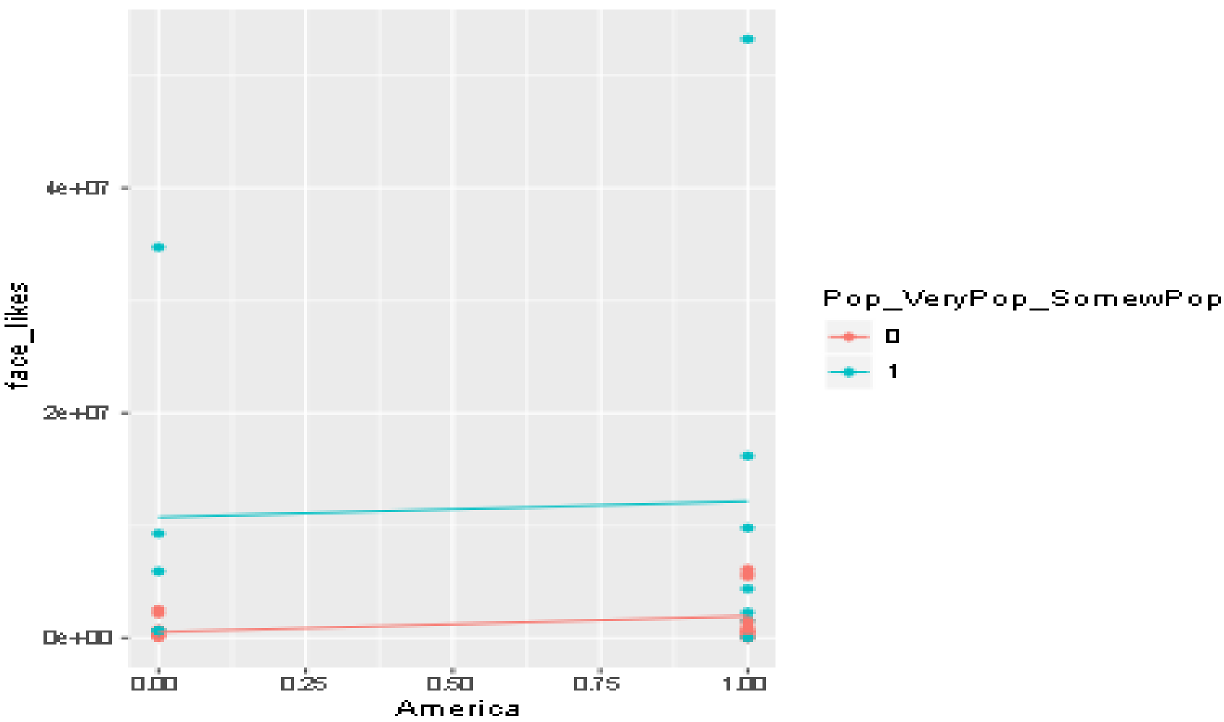
	face_likes	
	Model 1	**Model 2**
|America|	1427013.000	
	(3566917.000)	
|Europe and Central Asia|		-4514912.000
		(3539914.000)
Pop_VeryPop_SomewPop	10206996.000***	9941036.000***
	(3623092.000)	(3539914.000)
Constant	524281.200	3127637.000
	(3105342.000)	(2522885.000)
N	33	33
R-squared	0.211	0.248
Adj. R-squared	0.159	0.198
Residual Std. Error (df = 30)	10005852.000	9771091.000
F Statistic (df = 2; 30)	4.014**	4.938**

***p < .01; **p < .05; *p < .1

Lastly, we made two scatterplots in order to better interpret our two multiple regression models. Unfortunately, both scatterplots below don't show much association between the variables because the plots are not scattered. If they were, our predictor variable model could possibly be linearly predicted from the others with a substantial degree of accuracy and, therefore, we would have a significant collinearity value. In this sense, the two graphics show that our sample is not correlated. This means that our model is possibly biased or, as I prefer to imagine, the small number of cases don't show a substantive level of significance. One difference that we can observe is that the arrows of the two plots are diferent, and this can be explained by looking at the negative value to Europe and Asia value. More graphics may be seen in the methodological appendix



face_likes
4e+07
2e+07
0e+00
0.00
0.25
0.50
0.75
1.00
Europe_CentralAsia
Pop_VeryPop_SomewPop
0
1



face_likes
4e+07
2e+07
0e+00
0.00
0.25
0.50
0.75
1.00
America
Pop_VeryPop_SomewPop
0
1

6 IMPLICATIONS AND CONCLUSIONS

As noted in the previous section, the multiple regression tables and graphs do not indicate correlation between the variables and the model. Thus some measurement parameters were observed that differ from the standarts of significance and adjustments.

However, it is necessary to clarify that it is very likely that the results were not satisfactory (see also methodological appendix for more information) because the number of observations of our created Dataset is small. It is believed that the present research has potential for future studies to adapt the model, inserting more cases so that the regression analysis has more substance.

Indeed, the literature on the theme of social networks as a stage for the spread of populist discourses is still recent, with few theories about it, indicating that there is no robust studies and assumptions for the theme. Overall, this study aimed to find a correlation between the facebook media, populista discourse and geographical region.

We chose America and Europe And Central Asia variables because, specially in latin america, there’s a strong historical path with the presence of populista political. We highly encourage the academy to investigate thoroughly about the different implications of populism in social media, becase, as we have seen from the Team Populim ans The Guardian Research, the number od populista politics around the glob is increasing, and the information has something to do about it.

7 BIBLIOGRAPHY

BIMBER, Bruce, 1998. **The Internet and political transformation: Populism, community, and accelerated pluralism**. Polity, 31(1), 133–160.

CANOVAN, Margaret, 1999. '**Trust the People! Populism and the Two Faces of Democracy'**, Political Studies, 47 (1), 2–16. Available at https://journals.sagepub.com/doi/10.1111/1467-9248.00184. Accessed in August

HAWKINS, Kirk A., and SILVA, Bruno Castanho, 2018. **"Text Analysis: Big Data Approaches"**. In: The Ideational Approach to Populism: Theory, Method & Analysis, edited by Kirk A. Hawkins, Ryan Carlin, Levente Littvay, and Cristóbal Rovira Kaltwasser. London: Routledge.

LEWIS, Paul, CLARKE, Seán and BARR, Caelainn. '**How we combed leaders' speeches to gauge populist rise.** The Guardian Newspaper. London, 6 March 2019.Avalable at <https://www.theguardian.com/world/2019/mar/06/how-we-combed-leaders-speeches-to-gauge-populist-rise> Accessed in july, 2019

MOFFIT, Benjamin, 2016. **The Global Rise of Populism: Performance, Political Style, and Representation**, Stanford University Press. Stanford. 240 pgs.

SHILS, Edward, 1956. **The Torment of Secrecy: The Background and Consequences of American Security Policies**. New York: Wiley

WISHLADE, Fiona, 2019. **The Rise of Populism, Regional Disparities and the Regional Policy Response**. European Policy Research Paper. University of Strathclyde Publishing. Available at:<https://www.google.com/search?q=the+rise+of+populism%2C+regoional+disparities&oq=the+rise+of+populism%2C+regoional+disparities&aqs=chrome..69i57j69i64.12227j0j4&sourceid=chrome&ie=UTF-8>. Accessed in 8 of August, 2019.

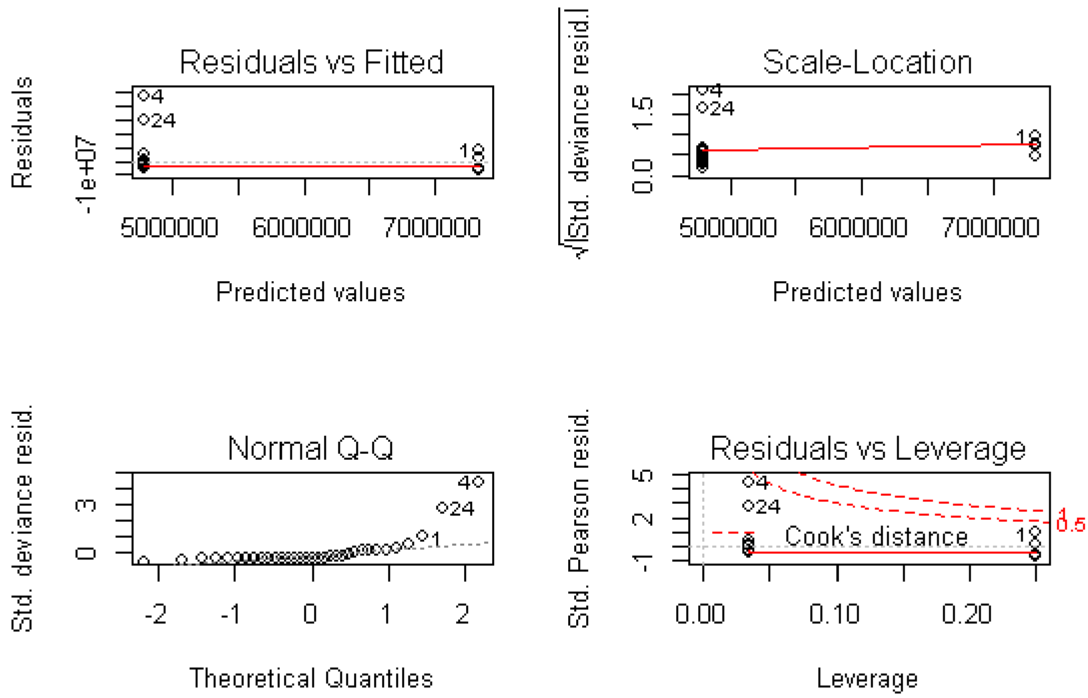
METHODOLOGICAL APPENDIX

The next five plots below shows all the regression results. When it concerns to the multiple linear regression analysis, it requires that the errors between observed and predicted values (i.e., the residuals of the regression) should be normally distributed. This assumption may be checked by looking at the Q-Q-Plot (Or Quantile-Plot). All Q-Q plots below are “heavy scaled”, which indicates a lot of concentration and that the distribution hasn’t the same shape. Another importante feature is homoscedasticity. This assumption states that the variance of error terms are similar across the values of the independent variables.

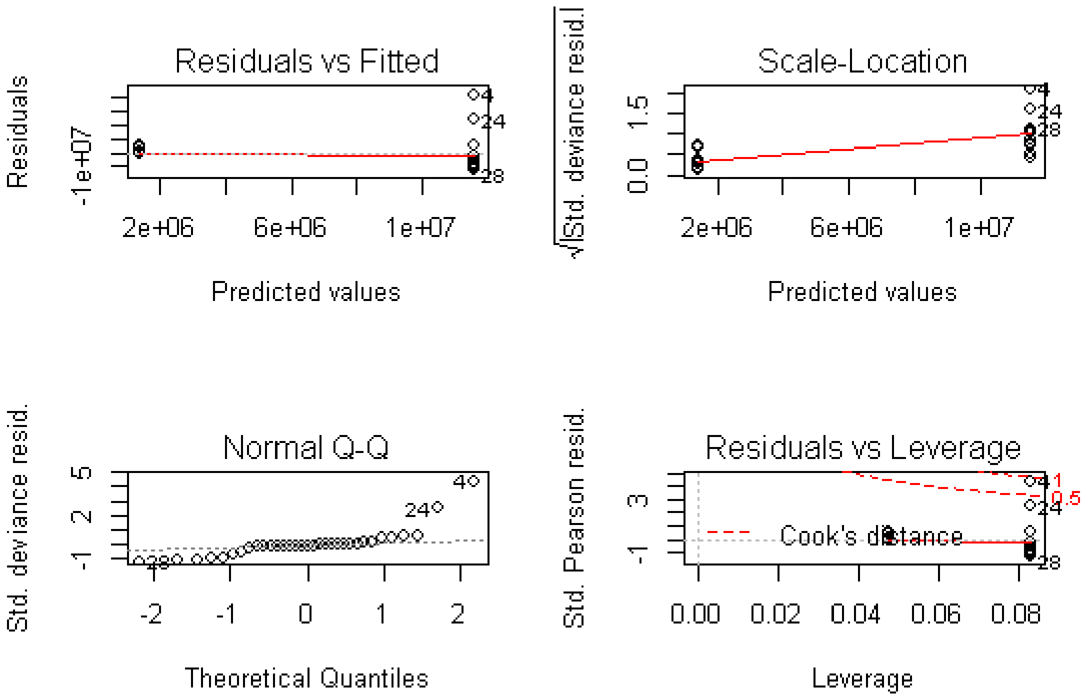
Moreover, a scatterplot of residuals versus predicted values is good way to check for homoscedasticity. There should be no clear pattern in the distribution; if there is a cone-shaped pattern (as shown below), the data is heteroscedastic.

Thus, the five graphs show some similarity in the way points are distributed. Thus, it is clear that the data are heteroscedastic; there are several outliers that move away from the line, which indicates that many observations are away from the adjusted model (and, therefore, the farther the value is, the more residual it contains.

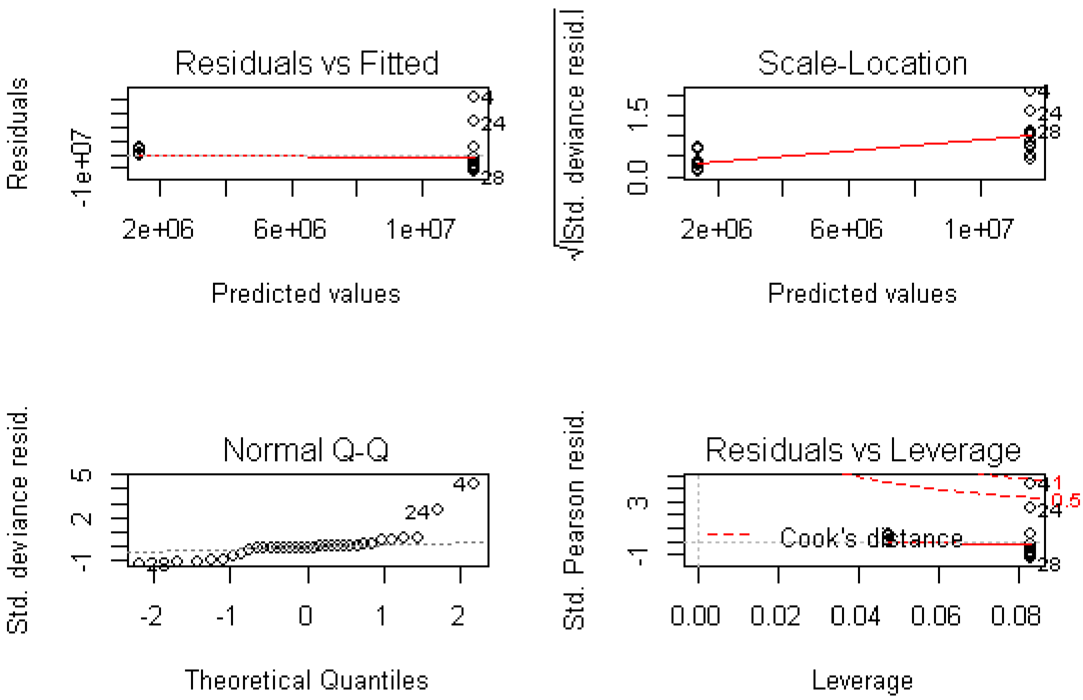
Reg1_Pop_VeryPop Results face_likes ~ all populist scores less “somewhat populist” and “not populist”)



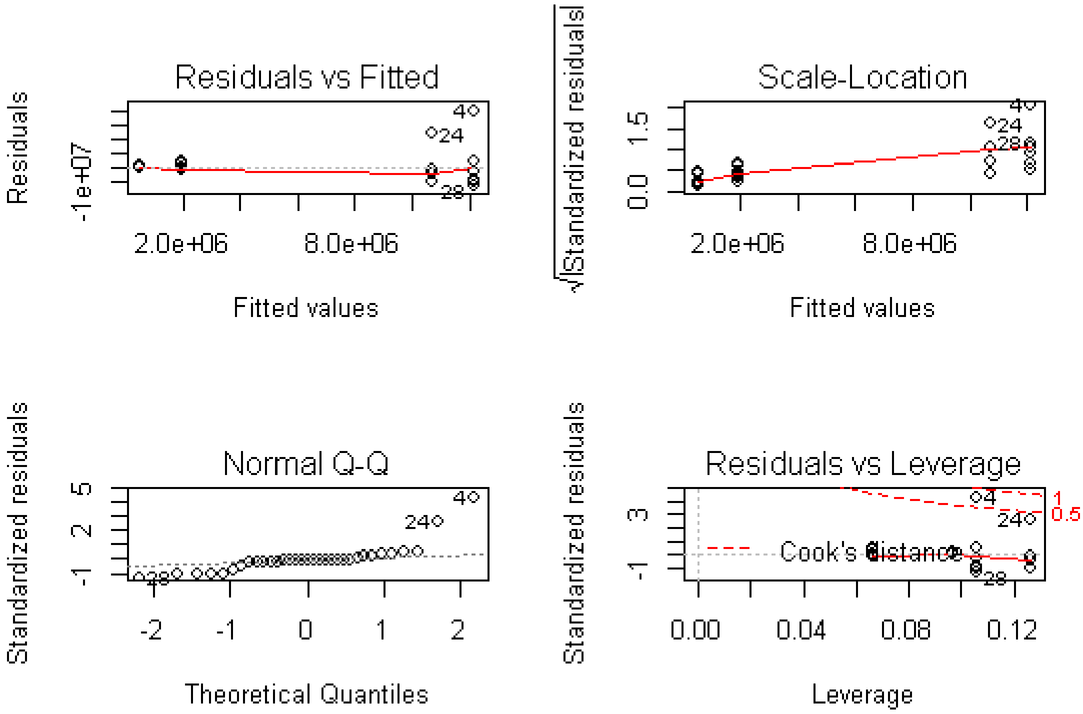
Reg2_with_Somewh (face_likes ~ all populist scores less not populist)



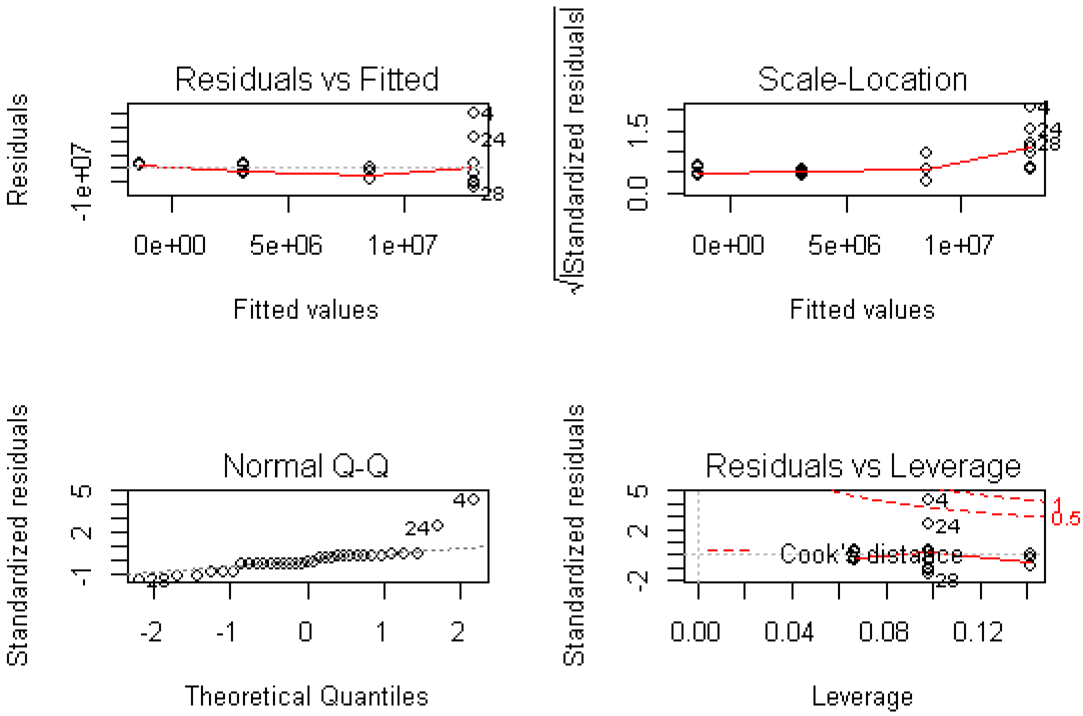
Reg3_Not_populist(face_likes ~ not populist)



RegM1_America (face_likes ~ all populist scores + america)

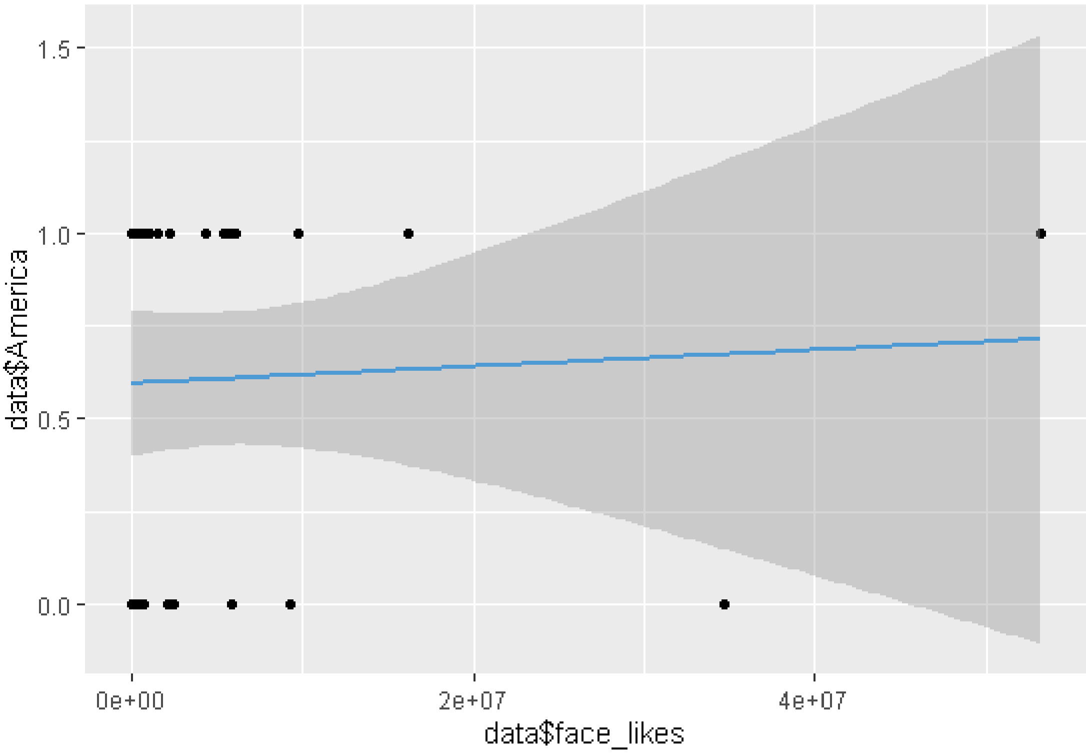


RegM2_Europe_CentralAsia(RegM1_America (face_likes ~ all populist scores +Europe_CentralAsia)

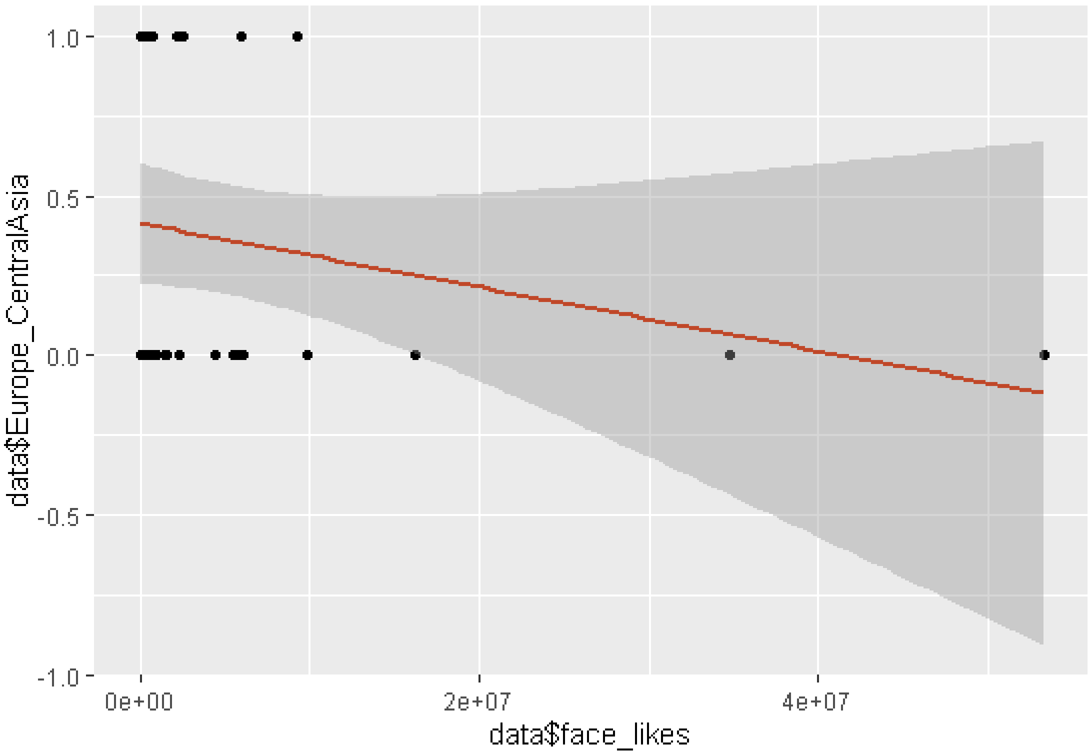


The last two graphics below show na alternative illustration of the scatterplot, but without interaction. As we have already examined, they have some problems, specially because the points are not randomnly scattered.

RegM2_Europe_CentralAsia(RegM1_America (face_likes ~ all populist scores +Europe_CentralAsia)



RegM2_Europe_CentralAsia(RegM1_America (face_likes ~ all populist scores +Europe_CentralAsia)


1.0
0.5
0.0
-0.5
-1.0
data$Europe_CentralAsia
0e+00
2e+07
4e+07
data$face_likes

~

Tratamento de Dados (tratamento.R)

1 Carregando Pacotes

```
if(require(stargazer) == F) install.packages('stargazer');
require(stargazer)

## Loading required package: stargazer

##
## Please cite as:

##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and
Summary Statistics Tables.

##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

library(stringi); library(stringr);
library(dplyr); library(readxl)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

2 Carregando Dados

```
face <- read_excel("dados/facebook_2019.xlsx")
face2 <- read_excel("dados/world_leaders_followers_atualizado.xlsx")

## New names:
## * `` -> ...1
## * `` -> ...2
## * `` -> ...3
## * `` -> ...4
## * `` -> ...5
## * ... and 24 more problems

pop_speech <- read.csv("dados/pop_speech_clean.csv")
```

3. filtrar selecionar na base do facebook

```
face = face[face$`Personal Profile` == 1,]
```

4 Tirar Nomes 'Prime Minister' e 'President'

```
face$leader = str_replace(face$`Page Section`, 'President ', '')
face$leader = str_replace(face$leader, 'Prime Minister ', '')
```

5 remover caracteres para padronizacao e combinacao

```
face$leader = stri_trans_general(face$leader, "latin-ascii")
face$leader = tolower(face$leader)
face$leader = str_replace(face$leader, '-', ' ')
```

6 Renomear colunas a partir da linha 2 e remover linha 1

```
colnames(face2) = face2[2,]
face2 = face2[-c(1:2),]
```

7 transformar likes em numerico

```
face2$Likes = as.numeric(face2$Likes)

## Warning: NAs introduzidos por coerÃ§Ã£o
```

8 remover caracteres para padronizacao e combinacao

```
face2$leader = stri_trans_general(face2$`World Leader`, "latin-ascii")
face2$leader = tolower(face2$leader)
face2$leader = str_replace(face2$leader, '-', ' ')
```

9 selecionar variaveis de interesse

```
face = face[,c('leader','Likes')]
face2 = face2[,c('leader','Likes')]
```

10 concatenar com a base facebook19

```
dataface = rbind(face, face2)
```

11 remover casos dusplicados e casos faltantes

```
dataface = dataface[!duplicated(dataface$leader),]
dataface = dataface[complete.cases(dataface$Likes),]
```

12 combinar dados

```
dataset = merge(pop_speech, dataface, by='leader')
```

13 selecionar variaveis e renomear

```
dataset = dataset[,c('leader', 'Likes','average.score',
'speech.category', 'region')]
colnames(dataset) = c('leader', 'face_likes','average_score' ,'speech',
'region')
```

14 Salvando a nova base em formato csv e xlsx

```
write.csv(dataset, 'dataset_cru.csv', row.names = F)
```