

Erasmus University Rotterdam

Erasmus School of Economics

MSc thesis program Behavioral Economics

*Does asking about social circles improves
election predictions in Colombia?*

Author: Nicolás Ramírez Saldarriaga

Student number: 597201

Supervisor: Cem Peker

Second assessor: Johan de Jong

Date: 02 August 2022

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics, or Erasmus University Rotterdam.

Abstract

Galesic et al. (2018), showed that asking about social circles in elections polls with two parties could reveal a better outcome. Later Bruine De Bruin et al. (2018) replicated the experiment with multiple parties in the Netherlands, again obtaining more accurate results. I replicated the experiment by running a poll for the multiple-party presidential elections that were held on the 29th of May 2022 in Colombia, specifically in Bogotá. Additionally, aiming to identify some characteristics of those with less reliable answers I asked the respondents their prediction of the results and calculated the prediction errors. The results showed that asking about social circle behavior provides more accurate results among people who voted in Bogotá. On the other hand, Colombians are classified into strata or socioeconomic groups according to their houses and area of residence. There are six social strata: strata 1 to 3 are beneficiaries of subsidies for the domestic public services, while strata 5 and 6 pay an additional contribution in their public services that is used to subsidy people from strata 1 to 3. Hidden poverty in strata 4 and 5 shows that they are no longer benefiting from the current status quo, therefore they might not have any interest in maintaining it and might give less reliable answers about their voting intentions. After the analysis, there was not enough evidence to state that strata 4 and 5 had the highest misreporting voting intentions in Bogotá.

Keywords: Social circle, election prediction, survey data, Colombia.

Table of Contents

1. Introduction	4
2. Literature Review	6
Potential discrepancy between self-reported preferences and actual voting behavior	6
Colombia's 118 th president	10
Hidden Poverty	12
A turn to the left	12
3. Methodology	15
Experimental design	15
Sample size calculation	16
4. Analysis	19
5. Results	23
Simple vote aggregation vs. social-circle estimate	23
Identifying less reliable	25
Comparing means and variances	25
6. Discussion	27
7. Conclusion	31
8. References	32
Appendix A	44
Survey	44
Consent	44
Personal questions	45
Social circle questions	47
Demographics	50
Final question results	51
Appendix B	54
Education level Bogotá	54
Appendix C	55
Appendix D	57

1. Introduction

This year Colombia is facing one of the most controversial presidential elections in its history. The atmosphere a week before the elections is so tense that the media is already talking about a possible coup d'état if the left party wins (Caracol Radio, 2022b; El Colombiano, 2022; El Espectador, 2022; and Publimetro Colombia, 2022). The tension increase as the election day is closer, and the polls point out that the left party candidate will be the winner (Celag 2022; Centro Nacional de Consultoria, 2022; Guarumo y Ecoanalitica, 2022; Invamer, 2022; and YanHass, 2022). But should we trust this survey data?

Some of the reasons why sometimes surveys are not able to predict the behavior of the respondents are: (1) not understanding or being confused about the question, (2) just a human error of choosing a different answer, (3) and consciously lying (Enns et al., 2017, Krumpal, 2011). Nevertheless, election polls are still important, according to Kuha (2022), some of the election polls' purposes are: (1) a feedback mechanism used by parties' policy choices on how the campaign is going; (2) and to help politicians understand what the public wants and views of specific groups of people.

Trying to get more precise predictions Galesic et al. (2018), showed that asking about social circles in elections polls with two parties could reveal a more accurate vote distribution. Later Bruine De Bruin et al. (2018) replicated the experiment with multiple parties in the Netherlands, again obtaining more accurate results. In this case, I am willing to verify if their findings could be replicated in Bogotá, Colombia. Therefore, my research question is:

Does asking about social circles improves predictions of national election results with many political parties in Bogotá, Colombia?

Another approach to gather more accurate vote distribution is by identifying those with less reliable answers. Not taking their answers into account could lead to more trusty results. This task is even harder in surveys where it is almost impossible to prove if the respondents are telling

the truth or not. By asking the respondents to predict the final results and compare them with the actual ones (prediction errors) I will try to identify some of the characteristics of those who could give less reliable answers. According to Luhmann (1996), mass media has an impact on the way we learn and see the world. Additionally, the media are assumed to have a potential for significant influence (McQuail, 2010). Therefore, the prediction results of each respondent should be broadly the same. To do this I will group Bogotá's population by socioeconomic classification and compare the prediction errors.

2. Literature Review

Hypothetical questions might induce hypothetical answers. This has to be taken into account when running surveys that include self-report questions. A slight change in the words, order, format, or context can affect the results (“A Step-by-Step Guide to Smart Business Experiments,” 2014, and Bertrand & Mullainathan, 2001). Additionally, questions about behavior may be viewed as threatening and may result in biased reports. Even voting questions might create some discomfort for the respondents as they could believe that the interviewer could judge them based on their answers (Bradburn et al., 2004). Moreover, behavioral questions rely on people’s abstraction of the general population and this inference can be influenced by external factors (Galesic et al., 2018).

To increase the accuracy of the polls, it is necessary that the respondent’s understanding of the question is as close as possible to what the researcher has in mind. Nevertheless, there are cases in which both, the researcher and the respondent’s understanding of the question is the same but the respondent decides to lie.

Potential discrepancy between self-reported preferences and actual voting behavior

The effectiveness of survey research depends on how honest people are reporting their beliefs and preferences. Unfortunately, there are several reasons why there might be differences between the results of polls and the actual behavior of the respondents. First, it might be due to an accidental misreporting, for example, Brosnan et al. (2019), by tracking the online respondents’ eyes, found that they spend just 32% of the needed time to carefully read the instructions, questions, and answer options of the surveys.

The misreporting can be caused simply by confusing and selecting the wrong answer or by the disinterest of the respondents in answering the survey. For instance when people just give a random response or do not think carefully before answering. This behavior tends to increase throughout the survey, primarily because respondents assume to be familiar with the content. Bowling et al. (2016) define this phenomenon as insufficient effort responding. Additionally,

Meade and Craig (2012) highlighted some of the causes of inattention such as: having a low level of motivation or interest, low personalization of the survey, the difficulty of the questions, physical distance from the investigator, and environmental distractions.

On top, respondents consciously might lie. For this research, I will use the definition of lying stated by Krauss (1981), “an act that is intended to foster in another person a belief or understanding which the deceiver considers to be false”. This definition specifically addresses the communicator's intention and differentiates false beliefs from lying. Despite this differentiation, deceiving others could be considered common, taking into account that on average we do it more than once a day (Vrij, 2008). However, DePaulo et al. (1996) showed that shown that the motivation for lying is less driven by monetary gain and material advantages and more by emotional rewards like esteem, affection, and respect. Additionally, Goffman (1959) showed that many of the lies told on a daily bases are said to avoid tensions or conflicts.

Additionally, respondents can lie in election polls because of social pressure. According to Asch (1995), people could adjust their judgments so they match the views of the majority of the experts. He also showed that the tendency to conform to the majority increases once the person is found alone (Asch, 1955). Based on these findings, the polls could create a pressure loop, therefore reducing their accuracy. As more poll results are released the decision of the majority becomes more clear, hence the respondents of the polls have more pressure to lie. Therefore, the closer it is to the election date, the higher the chances of a respondent lying in a poll.

Moreover, this situation might be worse in surveys where most of the questions focus on the individual self-reported and the respondents' behaviors. Helzer and Dunning (2012) argued that people’s self-predictions tend to be biased and overestimated because individuals could feel the need to base their choices on their expectations of successful future scenarios.

[How to improve data quality](#)

By acknowledging the potential discrepancy between self-reported preferences and actual behavior, but also the importance of understanding the preferences of humans, there are methods for improving the quality of the data. First, researchers should pay attention to the layout and distribution of the questions and answer options. For example, to avoid respondents losing interest or selecting the answer they are willing to give, Brosnan et al. (2019) recommend dividing the survey questions into different screens.

Another method is when researchers use scoring rules to incentivize respondents in a certain way that telling the truth is more profitable for respondents. One of the methods used is Prelec's (2004) Bayesian Truth Serum, which has been developed to incentivize unverifiable answers (Frank, Cebrian, Pickard, & Rahwan, 2017; Kukla-Gryz, Tyrowicz, Krawczyk, & Siwiński, 2015; Weaver, & Prelec, 2013). This method asks respondents to answer a question and to predict how common each possible answer will be. Respondents are rewarded for the accuracy of their predictions by having surprisingly common responses and penalizing those with surprisingly uncommon answers.

As well, asking respondents about the behavior of their social circles has shown to be a good alternative to increase the accuracy of the election polls (Galesic et al., 2018). For example, Nisbett and Kunda (1985) showed that estimations of social distribution characteristics made by people are close to the real outcome. This technique is based on the idea that respondents can be used as informants, as they can learn about the behavior of their social circle by observing them, or talking with them about it (Bradburn et al., 2004). Therefore, these type of questions provides information about individuals who were not included in the sample. Additionally, when respondents are asked about the behavior of unidentified friends or relatives they will not be so threatened as they would be answering the same question about their behavior. (Bradburn et al., 2004). Galesic et al. (2018) showed how the accuracy of the election polls can be improved, reducing the lies of the respondents due to social pressure, during the U.S.A presidential elections in 2016 and the 2017 French presidential elections.

Additionally, Williams & Gilovich (2008) found evidence that predictions based on the actions of friends and family (social circle) are typically less idealistic and more accurate in terms of the actual results. Therefore, as asking about the behavior of the respondents' social circle has been proven to lead to more precise results than self-prediction questions, it will help to reduce the aspirational bias and the strength of the respondents' intentions (Helzer et al, 2012).

Despite the success mentioned before, it is also important to state the limitations of this method. Media plays an important role in how respondents might inform themselves about what other voters will do. Irwin & van Holsteyn (2002) results showed that the more attention people paid to polls, the more their estimates tended to converge around the predictions stated by the polls. Even though, this method explicitly asks the respondents "What percentage of your social contacts are likely to vote in the upcoming election?" and "Of all your social contacts who are likely to vote, what percentage do you think will vote for [candidate]?" (Irwin & van Holsteyn, 2002). It is hard for respondents to differentiate their friend's behavior from what the media states, these two factors (media and social circle perception) can be mixed up easily. Therefore, there is a high probability they will not be answering what friends might do and instead what the polls say people might do.

The impact of media on people's behavior cannot be isolated. Media can also have an effect the importance people give to specific issues, based on the amount of times and information they receive of it (McCombs & Shaw, 1972). Media companies might use this knowledge to build up their agendas and arrange the importance they want to be set to each topic, this is called Agenda Setting (McCombs & Shaw, 1972). Cohen (1963) argued that the press might not be successful in telling people what to think, but impressively good at telling them what to think about. With the democratization of the internet, the power of Agenda Setting has changed and now big media companies have lost some of it, as consumers have become also content producers, and prosumers (Toffler, 1980; Ritzer et al., 2012). This shows that throughout history what people perceived to be more important can be shaped by the behavior of others and how they receive the information, either by what big media corporations decide to share with their viewers or

what other consumers have decided to share on the web. Therefore, when people are asked about the behavior of unidentified friends, they may answer not only based on what their friends do but also mix it with what they have seen on different platforms.

Colombia's 118th president

In 2022 Colombians will elect their 118th president. According to the 191st article of the current Colombian Constitution, the President of the country will be elected for four years, by half plus one of the votes. If no candidate obtains this majority, the two candidates who obtained the highest number of votes will move on to a second round. This second round will take place three weeks after the first one. The person who obtains the highest number of votes will be Colombia's next president (Colombia, 1991). During the first round of elections, an experiment will be conducted to see if the method of asking about social circles can improve election predictions in Bogotá. Based on the theory presented previously, the first hypothesis is

H1. Asking about social circles will improve predictions of national election results with multiple parties in Bogotá, Colombia.

The predictions shared for the last two country elections in Colombia have been controversial. In 2016 a referendum allowed Colombians to decide if they accepted or not the peace agreement that could be the end of the Colombian conflict between the Colombian government and the FARC guerillas. The opinion polls pointed out that 54% of the Colombians would accept the agreement; while 34% would not accept the agreement, and 12% did not know at the moment they were asked (Caracol Radio, 2016). The crushing results of the acceptance did not appear on the final results, and only 49.78% accepted while 50.21% voted against (Plebiscito 2 Octubre 2016 - República de Colombia, 2016). On the other hand, the election polls for the presidential elections in 2018 were very accurate and predicted that 57% of the Colombian voters were going to choose Ivan Duque as their new president, and 37% as Gustavo Petro (INVAMER, 2018). On the 17th of June 2018, Ivan Duque was announced as the 117th formal president of Colombia with 54% of the votes (Registraduría Nacional del Estado Civil, 2018).

Socio-economic stratification of people

We as humans are in steady got to build and keep up solid and steady interpersonal connections. (Baumeister & Leary, 1995). Being part of a group satisfies this need as it provides the members with assistance and social support. Although, for a collection of people to be a group there must be other people who are not in the group (Hogg, 2001). Additionally, there are cases where we are forced to be divided into groups without knowing it or requesting to be part of it. For example, based on our demographic qualities such as sex or age. In this sense, social groups are categories of people that acquire their meaning by contrast with other categories. No matter the reason we ended up being part of a group, some of the characteristics of a typical member of it will ascribe to us, and so stereotype ourselves (Hogg, 2001).

Another example of social categorization is the “socioeconomic stratification” in Colombia. This allows the government to classify the population into strata or socioeconomic groups based on their houses and where they are located. It was created to easily make sure that all Colombians have access to public services. As a result, those who live in better houses or areas (higher strata) will pay more for domestic public services, subsidizing those who live in houses or areas with lower living conditions (lower strata). Colombia has six social strata organized by number where: 1 is the lowest; 2 is the lower ; 3 is low; 4 is the medium; 5 is the higher, and 6 is the highest (DANE, 2022). People who live in strata 1–3 are beneficiaries of subsidies in their utility payments, paid by the additional contribution from those who live in strata 5–6. People in stratum 4 are the only ones who just pay exactly what they have consumed, they do not receive any subsidies.

This instrument was built based on statistical tests that showed that the characteristics of a house and area of residence were correlated with the living conditions of its inhabitants (DANE, 2022). According to DANE (2022), this relationship is caused because the characteristics of a house go beyond the basic need for shelter and more into psychological and esthetic aspects. Therefore, money is the main limitation for households to have a better house. According to law 142, article 101.1 (Colombia, 1994) socioeconomic stratification is a better way to group socially people

instead of by their income. The income per household is very difficult to calculate as it changes constantly due to informal working lifestyles and forced migration.

Hidden Poverty

According to the Internal Displacement Monitoring Center (IDMC) (2020), Colombia has the second highest number of internally displaced people in the world. Therefore, the justification to group people by their housing conditions might make sense. On the other hand, since the end of the '90s, the reduction of the middle class has increased significantly, reducing the purchasing power parity of those who lived in strata 4 and 5 (Secretaría Distrital de Planeación (SDP), 2013). Consequently, the statistical aggregation presents a different country than the one that could be seen in the currently available statistical data.

Trying to fix the loss in the statistical aggregation, Bogotá's local government defined hidden poverty when assumed parameters, rules, and structural correlations do not allow the government to detect poverty situations in higher strata (Secretaría Distrital de Planeación (SDP), 2013). This shows that there is a probability that the socioeconomic stratification number assigned to a house might mean that the households who live there should be considered poor or rich. Mendoza Torres et al. (2011) defined that a person should be considered poor if despite having a job or other sources of income, do not have sufficient resources to cover their needs, which include the basic ones and those referring to the standards of society. Not being able to cover the needs referred to the standards of society might have a greater impact in the higher strata, the reason why Mendoza Torres et al. (2011) renamed hidden poverty as shameful poverty (*pobreza vergonzante*).

A turn to the left

By 2008, eleven of the eighteen Latin American countries were governed by center-left or left-wing presidents (Stokes, 2009). Colombia was not part of those eleven countries, and until now has kept a right political tendency. This preference was influenced by the internal conflict between the government and some guerrilla groups like the Revolutionary Armed Forces of

Colombia (FARC), the National Liberation Army (ELN), and the Popular Liberation Army (EPL). According to the Permanent Committee for Human Rights (El Espectador, 2021), in Colombia since 1888 a series of laws have been established that legalized the crimes against those who militated in the communist party. According to Latin American Public Opinion Project (LAPOP Lab) and Vanderbilt University (2010), Colombia was one of the Latin American countries with fewer people proclaimed to be interested in a left political tendency.

By mixing the idea that respondents can lie or report unreliable answers in election polls because of social pressure; there is a group of people that are ashamed or embarrassed by being considered poor; and the idea that a left political tendency action might be considered illegal, the result is my second hypothesis:

H2. People who are assigned to strata 4 or 5 will report less reliable information about voting intention.

This hypothesis is based on the Spiral of Silence (Noelle-Neumann, 1974) that could have been created in Colombia, especially among people who live in strata 4 and 5. According to Noelle-Neumann (1974), the Spiral of Silence happens when individuals are aware that they have an unpopular opinion among the group they are part of and they decide to remain silent. Instead, if they notice that their opinion is also shared throughout the group they are more confident with sharing it. In this case, the popular opinion is the willingness to continue with a right-party government, reinforced by old laws that legalized the crimes against those who militated in the communist party (El Espectador, 2021). Additionally, according to Béné et al.(2014) people who have relatively large economic, financial or political interests in keeping the current status quo will apply certain resistance to change. Therefore, people who live in strata 4, 5, and 6 somehow in the past had a relatively large economic, financial or political interest in keeping the status quo of a right-party government. Hidden poverty in strata 4 and 5 shows that they are no longer benefiting from the current status quo, therefore they might not have any interest in maintaining the current status quo. Consequently, when people that are assigned to strata 4 and 5 will be

asked about their prediction of the final results (percentage of votes each candidate will have), they will give a less reliable answer compared to the actual results.

3. Methodology

A survey was used as the main instrument in trying to understand if asking about social circles improves predictions of national election results with many political parties in Bogotá, Colombia. The survey followed Colombian law 1581 of 2012 (Colombia, 2012) which states that all people have the right to know, update, and rectify the information that has been collected about them in a database or files. Therefore, before starting answering the survey each person was informed of the goal of the research and had to read and agree to the data collection process. Additionally, no names, emails, phone numbers, or personal information of the respondents was asked or stored, and even the IP address was disabled, assuring complete anonymity of the data collection process.

Experimental design

An online survey was created and sent to friends and family through social media. As the target population is people living in Bogotá, Colombia, the questionnaire was structured entirely in Spanish. The translations in English can be found in Appendix A. The survey was created taking into account the recommendations given by Brosnan et al. (2019), and almost every question was on a different screen, except for the demographic questions as all of them were on one screen. The structure of the survey was: first respondents needed to read and accept the consent, here they were informed about the goal of the survey and the anonymity of the answers. Second, they were asked about their behavior (if they were planning on voting and for whom). Third, they were asked about their social circle behavior (the percentage that was planning on voting and the percentage that they thought were going to vote for each candidate). Fourth, demographic questions (the city where they lived, socioeconomic number, gender, age, and the highest level of education achieved). Finally, they were asked to predict the results of the elections. In all the three cases when the respondents had to see the list of candidates, the order in which the candidates were presented was randomized.

The structure of this survey was not exactly the same as the ones designed by Galesic et al., (2018) and Bruine De Bruin et al. (2022). Almost all the questions were structured and used the

same words as Galesic et al., (2018) and Bruine De Bruin et al. (2022) did in their papers, except for the own intention questions. Both authors kept the structure of asking the percent chance of voting and the percent chance of voting for each of the candidates. Instead, I decided to simplify the question and asked straight forward if they were planning to vote and for whom they were planning to vote. Bruine De Bruin et al. (2018) argue that according to Bresser and Soest (2018) this probabilistic own-intention question was better at foreseeing participants' true voting behavior, compared to when they were asked using the own-intention wording question. Nevertheless, there is a difference in the education level between Europe and Bogotá. In 2020, 32.8% of the EU population between the ages of 25 and 64 years had a tertiary level of education, while 46.2% had completed upper secondary or post-secondary, not taking into account those who have a tertiary education (Eurostat, 2021). By contrast, in Bogotá, 60% of the population has not finished high school (DANE, 2018). The lack of education among people who live in Bogotá will make it harder for them to understand probabilities and assign a number to each case. Following the recommendations that were given by Brosnan et al. (2019), to make the surveys as simple as possible, the intention question was changed.

The link to the survey was shared on social media and with friends and family members nine days before the election day, and the respondents had until the night before the election day to answer the survey. To increase the number of respondents, I also asked my friends and family members to share the survey with their social circle, and so forth.

Sample size calculation

Some factors affect the decision when selecting a sample size. Israel (1992) presented some strategies for determining the sample size.

1. Using a census for a small population: requires using the entire population as a sample size as it eliminates sampling errors and provides data on all individuals in the population. In this case, it is geographically impossible and costly, and inefficient to target Bogotá's entire population that is willing and registered to vote. Therefore, this option was opted out.

2. Using a sample size of a similar study: each election year election polls are done in Colombia, consequently, there is previous information that can be used. Table 1 summarizes the last polls done in the country with the sample size used. Then, the sample size for this task should be higher than 182 and could be smaller than 505 complete answers.

Table 1

The sample size was used in different election polls.

Survey	Size sample
YanHaas (2022)	182
Centro Nacional de Consultoria (2022)	489
Celag (2022)	505
Guarumo y Ecoanalitica (2022)	286
Invamer (2022)	362

3. Using publish tables: Israel (1992) provides a table with the sample size for a given set of criteria. Table 2 shows the sample size requires for a population size bigger than 100,000 and according to different precision levels and a 95% confidence level. Therefore, the sample size for this task should be between 100 and 1,111 complete answers.

Table 2

The sample size was used in different election polls.

Precision Levels	Size sample
$\pm 3\%$	1,111
$\pm 5\%$	400
$\pm 7\%$	204
$\pm 10\%$	100

4. Using formulas: Israel (1992) presents two option formulas that could be used to calculate the sample size: Cochran (1963) and Yamane (1973). Additionally, he shows that the answers presented by these two options are very close together. Furthermore, Uakarn et

al. (2021) showed that for similar cases to this one the sample size determination formula (1) by Yamane (1973) is suitable for survey research, taking into account that the formula has a 95% confidence level.

$$n = \frac{N}{1 + Ne^2} \quad (1)$$

Where:

- n is the sample size.
- N is the population size. According to Registraduría Nacional del Estado Civil (2022) there are 5,938,976 people allowed to vote in Bogotá.
- e is the desired level of precision.

Therefore, Table 3 shows the results of the formula by replacing the values and changing the desired level of precision to 1%, 5%, and 10%

Table 3

The sample size required according to each level of precision.

Level of precision	Sample size
1%	9,983
5%	400
10%	100

Based on this information, the sample required for this task should be between 100 and 9,983, depending on the level of precision desired. Precision levels of 1% and 3% are very high and costly inefficient, therefore the target number of answers will be between 100 and 400.

4. Analysis

To test if asking about social circles will improve predictions of national election results with multiple parties in Bogotá, Colombia (H1), the information was analyzed similarly to what Galesic et al., (2018) did in their paper, although some extra calculations were done. Two types of results will be presented for the individual responses and the social circle. For the individual values, first, vote share will be calculated (Ind Vote Share). The answers to the 3rd question (for who were they planning to vote?) were added to get the sum of the total votes, 319. Then the percentage of each candidate was calculated by dividing the number of voters registered by 319 and the number of votes that candidate had; the answer was multiplied by 100 to show it as a percentage. This process will allow me to measure the weight of each candidate from the total answers registered and extrapolate the percentages to represent the possible final results in Bogota.

The second result is a weighted vote share (Ind Wei Share) taking into account the distribution of the population in each socioeconomic number. By 2017 the distribution was: 1, 9,15%; 2, 41,37%; 3, 35,52%; 4, 9,42%; 5, 2,99%; 6, 1,55% (Diario La República, 2017). The weighting was done to correct for underrepresentation based on the strata. In this case, the answers were not only divided by the candidate but also by the strata mentioned by the respondent. Therefore, an extra step is added to the previous procedure. After, each candidate's vote for each stratum was weighted according to the distribution of the city. For example, 46 out of the 319 respondents said that they belonged to strata 2. Because 41,37% of Bogota's population is from strata 2 then the weight of these 46 respondents should also have the same weight; the 41.37% of 319 is 131,97. By cross-multiplication, each candidate's vote from a respondent from strata 2 can be transformed to match the weight of the city's population distribution. For example, if 6 respondents that belong to strata 2 mentioned they were planning to vote for candidate X. The weighted result for candidate X from people from strata 2 is $(6 \times 131,97) / 46 = 17,31$. Once this was done for each stratum for all the candidates then I ended up with a new list of votes for each candidate and then repeated the process described for the individual arithmetic average.

In line with Galesic et al. (2018) and Bruine De Bruin et al. (2018) papers, the prediction of election outcomes from social circle questions were obtained by (1) multiplying the percentage of each respondent's social circle that is likely to vote on the next elections by the percentage assigned to each candidate in the social circle question. (2) Once this is done, it is added up to next calculate the ratio and the average of the participant's social circle's likelihood to vote across all participants. This same process was used in this case. The participant's social circle's likelihood to vote was asked in the 4th question of the survey. The same questions used by Galesic et al. (2018) and Bruine De Bruin et al. (2018) were used for this experiment. The 4th question stated:

Think of your friends, family, colleagues, and other acquaintances 18 years of age or older that you have communicated with at least briefly within the last month, either face-to-face or otherwise. We will call these people your social contacts

What percentage of your social contacts is likely to vote in the upcoming election for President? For instance, 0% means that you think none of your social contacts will vote, and 100% means that all of your social contacts will vote. If you are not sure, just try to give your best guess.

Therefore, the answer given to this question was then multiplied by the percentage of the social circle that was planning to vote for each candidate (question 5). Then the result of this multiplication was divided by 100 (cross-multiplication). Afterward, the same process mentioned for the individual responses was done with this new dataset. For the vote share (SC Vote Share), all the votes were added up and used as the total, then each candidate's percentage was calculated. Additionally, for the weighted vote share (SC Wei Share), the same strata percentage distribution was used to calculate the weights.

For example:

- One respondent could have said that 80% of his/hers social circle is likely to vote in the upcoming election.
- Then he/she mentions that 40% of his/hers social circle will vote for candidate X
- To find the number of votes that represent the 40% out of the 80% that might vote the

following cross multiplication was done: $\frac{80}{100} = \frac{40}{x} \rightarrow x = \frac{80 \times 40}{100} \rightarrow x = 32$

- After doing the cross multiplication, all the new votes registered for each candidate (32 for candidate x in this example) were added together.
- This led to a new data set that with all the possible votes each candidate might have had. The same process to calculate the individual variables was done but with this new dataset.

The two same procedures used by Galesic et al., (2018) were used to compare the results and being able to establish which one is better. Both were presented by Mosteller et al. (1949) and are:

- Mosteller 3: is obtained by calculating the absolute error of the predicted vote share for each candidate and then computing an average of these values.
- Mosteller 5: is the difference between two differences of the two main candidates. First, is the difference between the percentages of the results of the survey of two candidates. Second is the difference between the percentages of the actual results of the same two candidates. The difference between these two results is the final value. For this error the two candidates chosen were Gustavo Francisco Petro Urrego – Coalición Pacto Histórico and Federico Andrés Gutiérrez Zuluaga - Coalición Equipo por Colombia. These two were chosen because when the survey was launched where the two candidates led the polls (Caracol Radio, 2022; Celag, 2022; Bloomberg, 2022; El Tiempo, 2022; Razón Pública, 2022 & Valora Analitik, 2022).

On the other hand, to measure the reliability of each respondent, the answer to the last question was compared to the actual results of the elections. As the survey was sent and answered online, it can be stated that none of the respondents is isolated from the media and therefore have had access to the same information. According to Luhmann (1996), mass media has an impact on the way we learn and see the world. Additionally, the media are assumed to have a potential for significant influence (McQuail, 2010). Therefore, the main idea of the final results of the elections each respondent had should be broadly the same.

Each respondent's prediction for each candidate was compared to the actual results, and the prediction errors were calculated. The protests that started on the 28th of April 2021 in Colombia showed that the socioeconomic numbers were more than a way to divide the population socioeconomically and more a way to identify the cultural, aesthetic, and political profile of a person (BBC News Mundo, 2021). Therefore, to verify if the prediction errors grouped by strata are different from each other a Kruskal Wallis test will be used. This test will show if the median of the 6 different samples (strata) comes from the same population. This test is used because there is no order in the groups. Therefore, the hypothesis will be

$$H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = \theta_6$$

$$H_1: \theta_i \neq \theta_j \text{ for some groups } i \text{ and } j$$

where θ_i is the median of group i .

5. Results

On the 28th of May 2022, the survey was closed with 330 answers. The analysis will only be conducted with the information provided by 319 respondents who finished the whole survey. The answers from 2 respondents were not considered because they did not agree with the term and conditions shown at the beginning. Furthermore, 9 respondents answered that they did not have plans on voting on the 29th of May, so their answers were not analyzed. In Appendix C are the results for all the descriptive variables.

Simple vote aggregation vs. social-circle estimate

Table 4 shows the results of the elections (Registraduría Nacional del Estado Civil, 2022) and the ones obtained in the survey.

Table 4

Comparison of the results of actual elections and polls.

Candidate	Actual Results	Indi Vote Share	Indi Wei Share	SC Vote Share	SC Wei Share
Rodolfo Hernández Suárez - Movimiento Liga de Gobernantes Anticorrupción	22,15%	18,18%	29,95%	14,70%	24,77%
John Milton Rodríguez González - Partido Colombia Justa Libres	1,25%	1,57%	3,86%	0,91%	2,23%
Federico Andrés Gutiérrez Zuluaga - Coalición Equipo por Colombia	19,24%	49,53%	24,35%	50,51%	27,27%
Sergio Fajardo Valderrama - Coalición Centro Esperanza	7,95%	12,23%	7,93%	8,75%	7,66%

Enrique Gómez Martínez - Partido Movimiento de Salvación Nacional	0,26%	1,57%	2,55%	1%	1,55%
Gustavo Francisco Petro Urrego – Coalición Pacto Histórico	47,05%	12,23%	18,85%	20,15%	29,31%
Ingrid Betancourt Pulido - Partido Verde Oxígeno	0,06%	0,31%	0,90%	0,53%	1,15%
Voto en blanco	1,93%	4,39%	11,62%	3,43%	6,07%
Error Mosteller 5 (1)		65,11%	33,31%	58,17%	25,77%
Error Mosteller 5 (2)		30,85%	36%	19,45%	20,36%
Error Mosteller 3		9,71%	7,07%	8,68%	4,52%

As it happened in 2016, again, the polls fail to predict the final results (Pulzo, 2022). Instead of Federico Andrés Gutiérrez Zuluaga, Rodolfo Hernández Suárez was in second place and now has the option to be Colombia's next president. None of the polls predicted this situation, not even this experiment's poll. This created the need that two Error Mosteller 5 were calculated. As mentioned in the previous chapter, the Error Mosteller 5 was going to be calculated comparing the two strongest candidates, based on the polls, but the results showed a different combination. Therefore, Error Mosteller 5 (1) is between Gustavo Francisco Petro Urrego and Federico Andrés Gutiérrez Zuluaga; and Error Mosteller 5 (2) is between Gustavo Francisco Petro Urrego and Rodolfo Hernández Suárez.

The results in Table 4 show that asking about social circles improves predictions of national election results with multiple parties in Bogotá, Colombia (H1). The Social Circle Voted Share (SC Wei Share) has the lowest percentage of errors compared to the actual results (in two out of the three different ways the error was calculated). Additionally, when just comparing the vote shares, also the social circle has the lowest percentage errors. Finally, when comparing the results of the social circle questions (vote share, vs weighted vote share) shows that the weighted vote share

is more accurate in two out of the three. Just, in the Error Mosteller 5 (2) the social circle vote share had a lower value than the social circle weighted vote share. These results support the idea that socioeconomic number is more than a way to divide the population socioeconomically and more a way to identify the cultural, aesthetic, and political profile of a person (BBC News Mundo, 2021).

Identifying less reliable

Comparing median

After calculating the prediction error for each respondent, the Kruskal–Wallis test was run, with the following hypotheses:

$$H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = \theta_6$$

$$H_1: \theta_i \neq \theta_j \text{ for some groups } i \text{ and } j$$

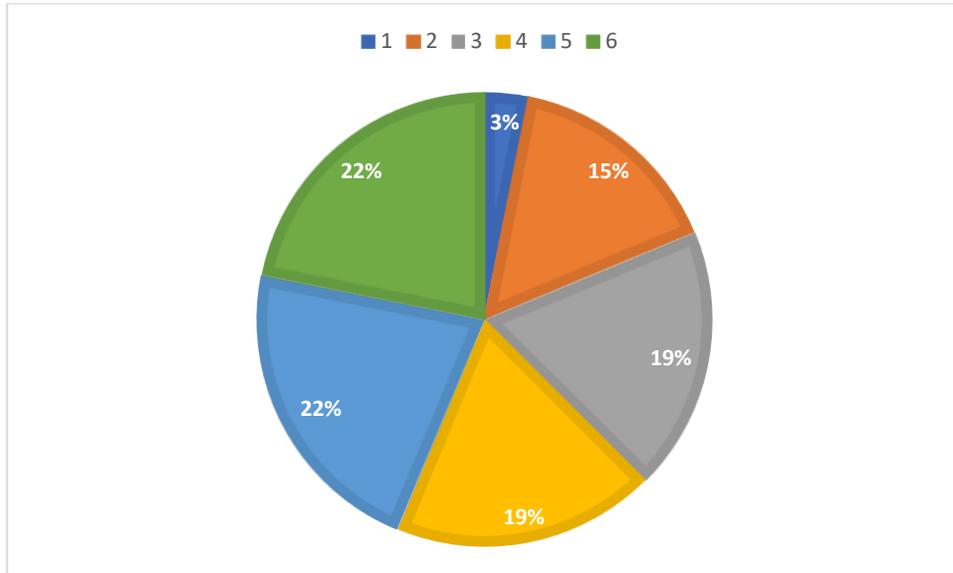
where θ_i is the median of group i .

The result was a p-value= 0,24 Consequently, it is not possible to reject the null hypothesis (H_0) at a 10% significance level. This result shows that there is no statistically significant evidence that at least two medians are different. Therefore, by using this approach it cannot be stated that respondents that are part of strata 4 and 5 in Bogota are more likely to give less reliable answers about their voting intention (H_2).

Additionally, the prediction errors were organized from highest to lowest and the 10% with the highest prediction errors were analyzed. Graph 1 presents the distribution of the strata of the respondents that were part of this group. This new group is divided almost equally between the strata. Despite the small difference between them, the majority belong to strata 5 and 6, followed by 3 and 4. This approach does not give any evidence that strata 4 and 5 are more likely to give less reliable answers about their voting intention (H_2).

Graph 1

Distribution by strata of the respondents with the highest prediction errors.



6. Discussion

After analyzing the data the two main limitations of this study are:

1. Fail to randomize the sample: although, according to the formula, the sample size was good (an error level of precision between 5% and 10%), and in the range of the previous polls done in the city, the respondents do not represent Bogotá's population. (1) The majority of the respondents are assigned to a socioeconomic number of 4 while the majority of the population of the city is 2. Additionally, in the weighted average, just one answer weighted 9.15%, as only one respondent marked being assigned to strata 1. (2) The education level in Bogotá follows the pattern as the country, presented in Appendix B. However, the distribution of the variable *Highest educational level achieved* does not follow this pattern. The popular answer for the highest educational level achieved was bachelor's (39,7%), while the majority of the population has barely finished primary. This might have happened because of the way the survey was distributed. As it was shared among my social circle, most people live in strata between 4 and 6. There was no significant participation of the lower strata who are the majority of the population.
2. Apparent lack of interest in the survey: as mentioned in the literature review, one of the reasons why we cannot trust survey data is because people are not motivated or interested in revealing the truth. 31 out of the 33 respondents with the highest prediction errors market either that 100% of the members of their social circle were planning on voting for one of the candidates, or that one candidate will have 100% of the votes. By basic understanding of probabilities, these two options are almost impossible, and this can be considered enough evidence to state that there is some lack of interest in answering the survey. Consequently, their answers can be considered less reliable. For that reason, the rest of the database was analyzed finding out that in total 83 respondents marked that 100% of the members of their social circle were planning on voting for one of the candidates. At the same time, 27 respondents expected that one candidate will have 100% of the votes.

By comparing these two groups I found that 88 respondents assigned, in at least one of the probability questions, 100% to just one candidate. As mentioned before, it can be stated that these 88 respondents gave less reliable answers. A new database was constructed after erasing the answers from these 88 respondents. The tables in Appendix D show the new distribution of each variable once the answers were erased. Additionally, it compares these new percentage distributions with the ones obtained during the previous analysis, with all the 319 answers. The order and distribution of the variables did not change drastically.

The data was recalculated only taking into account these 229 answers and the results are in Table 5. Based on the sample size calculation done before, this analysis will still have an error level of precision of around 5% and 10%. Additionally, it would be inside the range of sample size of the studies already done in Colombia. In this case, all the three errors calculated still show that the social circle questions predicted a more accurate outcome. The interesting aspect to point out is that the Social Circle Weighted Average was the only method to predict that the correct two top candidates in the correct order (Gustavo Francisco Petro Urrego – Coalición Pacto Histórico and Rodolfo Hernández Suárez - Movimiento Liga de Gobernantes Anticorrupción).

Table 5

Comparison of the results of actual elections and polls without untrustworthy answer.

Candidate	Actual Results	Indi Vote Share	Indi Wei Share	SC Vote Share	SC Wei Share
Rodolfo Hernández Suárez - Movimiento Liga de Gobernantes Anticorrupción	22,15%	20,52%	32,42%	17,62%	25,82%
John Milton Rodríguez González - Partido Colombia Justa Libres	1,25%	2,18%	4,83%	1,27%	2,68%

Federico Andrés Gutiérrez Zuluaga - Coalición Equipo por Colombia	19,24%	43,67%	23,31%	41,82%	24,89%
Sergio Fajardo Valderrama - Coalición Centro Esperanza	7,95%	14,85%	9,41%	10,77%	9,09%
Enrique Gómez Martínez - Partido Movimiento de Salvación Nacional	0,26%	1,75%	2,35%	1,37%	1,82%
Gustavo Francisco Petro Urrego – Coalición Pacto Histórico	47,05%	11,35%	15,71%	22,32%	28,09%
Ingrid Betancourt Pulido - Partido Verde Oxígeno	0,06%	0,44%	1,06%	0,74%	1,38%
Voto en blanco	1,93%	5,24%	10,91%	4,08%	6,24%
Error Mosteller 5 (1)		60,13%	35,41%	47,31%	24,61%
Error Mosteller 5 (2)		34,07%	41,61%	20,20%	22,63%
Error Mosteller 3		9,35%	7,85%	7,33%	4,76%

Based on my research, this was the first time this study was done in Latin America. Additionally, Ciocca and Delgado (2017), hold responsible for the lack of scientific productivity in America Latina to the political and economic instability. In the future, this study can be replicated with a bigger budget the problem of randomization and sample size can be solved. For this new experiment, I would suggest a phone survey instead of an online one to tackle the lack of probability understanding and increase the interest of the respondents.

As mentioned before, some of the election polls' purposes are: (1) a feedback mechanism used by parties' policy choices on how the campaign is going; (2) and to help politicians understand what the public wants and the views of specific groups of people (Kuha, 2022). On the other hand, people might change their opinions towards the views of the majority of the experts (Asch,

1955). Another study that could be carried on is to follow some people's voting elections through time and compare it with the news and other election polls' and final results. There is still a probability that people did not lie when answering the poll but the final results are different because they changed their minds in the meantime. This study can give a better understanding of how people change their minds and what factors have a higher impact on that. Once the extra factors are identified a model could be created to adjust the polls' answers based on these factors and increase the accuracy of the polls.

As mentioned in the Literature Review, one way to incentive respondents, to tell the truth, is Prelec's (2004) Bayesian Truth Serum. This method could be used to increase the accuracy of the election polls in Colombia. It will also allow spotting potential liars. This experiment will require respondents to answer the probability they have to vote for each candidate and predict the probability others have to vote for the same candidate. Additionally, this method requires an incentive to motivate respondents to tell the truth, for example, money.

7. Conclusion

This study showed evidence that asking people about social circle preference can improve presidential election with multiple parties in Bogotá, Colombia. Even though, a study with a better randomization sample is suggested to verify the implications and limitations of the hypothesis. More accurate results could give the politicians a better understanding of what the public wants and needs. For example, in this case, when the individuals that were identified as possible liars or with a lack of interest in the surveys were removed, the numbers revealed a potential interest in Rodolfo Hernández Suárez, that the other polls were not showing. Gustavo Francisco Petro Urrego's campaign focused on how to beat Federico Andrés Gutiérrez Zuluaga as the polls presented him as his biggest rival. Now that he is facing a different opponent he had to change his strategy affecting the credibility and efficiency of the campaign (Agencia EFE, 2022; La Patria, 2022; and, Vargas, 2022).

Despite that for this case there was not enough evidence to support the idea that people who belong to strata 4 and 5 would give less reliable answers for their voting intentions, there is still a possibility that there is social pressure on people who are part of the hidden poverty category. As Kuha (2022) mentioned, surveys are used to help politicians understand what the public wants and the views of specific groups. As mentioned before, Colombia is a country strictly divided into strata but surveys must reveal the truth of all the respondents so politicians can create more efficient campaigns and understand their specific needs. There might be some social pressure created by asking the strata in the election polls that could affect the understanding of the needs of those who are categorized as hidden poor.

8. References

- A Step-by-Step Guide to Smart Business Experiments. (2014, August 1). Harvard Business Review. <https://hbr.org/2011/03/a-step-by-step-guide-to-smart-business-experiments>
- Asch, S. E. (1955). Opinions and Social Pressure. *Scientific American*, 193(5), 31–35. <https://doi.org/10.1038/scientificamerican1155-31>
- Agencia EFE. (2022, June 9). La guerra sucia sale a flote en la campaña presidencial colombiana. Retrieved June 13, 2022, from <https://www.efe.com/efe/america/politica/la-guerra-sucia-sale-a-flote-en-campana-presidencial-colombiana/20000035-4828080>
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497–529. <https://doi.org/10.1037/0033-2909.117.3.497>
- BBC News Mundo. (2021, May 28). Qué son los estratos, el sistema “solidario” que terminó profundizando el clasismo y la desigualdad en Colombia. Retrieved April 26, 2022, from <https://www.bbc.com/mundo/noticias-america-latina-57264176>
- Béné, C., Newsham, A., Davies, M., Ulrichs, M., & Godfrey-Wood, R. (2014). REVIEW ARTICLE: RESILIENCE, POVERTY AND DEVELOPMENT. *Journal of International Development*, 26(5), 598–623. <https://doi.org/10.1002/jid.2992>
- Bergstrom, J., & Stoll, J. (1989). Application of Experimental Economics Concepts and Precepts to CVM Field Survey Procedures. *Western Journal of Agricultural Economics*, 14(1), 98–109.
- Bertrand, M., & Mullainathan, S. (2001). Do People Mean What They Say? Implications For Subjective Survey Data. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.260131>

Bickel, J. E. (2007). Some Comparisons among Quadratic, Spherical, and Logarithmic Scoring Rules. *Decision Analysis*, 4(2), 49–65. <https://doi.org/10.1287/deca.1070.0089>

Bloomberg. (2022, April 2). Encuesta revela que se cierra brecha entre Petro y Fico para la Presidencia colombiana. Bloomberg Línea. Retrieved April 26, 2022, from <https://www.bloomberglinea.com/2022/04/02/encuesta-colombiana-evidencia-mas-competencia-en-presidenciales/>

Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking Questions: The Definitive Guide to Questionnaire Design -- For Market Research, Political Polls, and Social and Health Questionnaires* (2nd ed.). Jossey-Bass.

Bresser, J., & Soest, A. (2018). The predictive power of subjective probabilities: probabilistic and deterministic polling in the Dutch 2017 election. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2), 443–466. <https://doi.org/10.1111/rssa.12409>

Brosnan, K., Babakhani, N., & Dolnicar, S. (2019). “I know what you’re going to ask me” Why respondents don’t read survey questions. *International Journal of Market Research*, 61(4), 366–379. <https://doi.org/10.1177/1470785318821025>

Bruine De Bruin, W., Galesic, M., Bååth, R., de Bresser, J., Hall, L., Johansson, P., Strandberg, T., & van Soest, A. (2022). Asking about Social Circles Improves Election Predictions Even with Many Political Parties. *International Journal of Public Opinion Research*, 34(1). <https://doi.org/10.1093/ijpor/edac006>

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology*, 111(2), 218–229. <https://doi.org/10.1037/pspp0000085>

Caracol Radio. (2022b, May 11). “Ha iniciado el golpe de estado en Colombia”: Quintero sobre suspensión. Retrieved May 19, 2022, from https://caracol.com.co/radio/2022/05/11/politica/1652232840_659137.html

Caracol Radio. (2022, April 9). Petro sigue con la preferencia de encuestados, según el Meta-Análisis de Infométrika. Retrieved April 26, 2022, from https://caracol.com.co/radio/2022/04/08/politica/1649453390_016033.html

Caracol Radio (2016, September 23). El Sí en el plebiscito lograría el 54% según encuesta sobre intención de voto. Caracol Radio. Retrieved January 26, 2022, from https://caracol.com.co/programa/2016/09/23/6am_hoy_por_hoy/1474630703_980326.html

CBS. (2018, September 6). Education - Figures - Society. Trends in the Netherlands 2018. Retrieved May 18, 2022, from <https://longreads.cbs.nl/trends18-eng/society/figures/education/>

Celag. (2022, April 21). Encuesta Colombia, abril 2022. Retrieved April 26, 2022, from <https://www.celag.org/encuesta-colombia-abril-2022/>

Centro Nacional de Consultoría. (2022, May 13). Intención de voto Línea de Base. Portafolio. Retrieved May 18, 2022, from https://www.portafolio.co/uploads/files/2022/05/15/ENCINA_SAS-LineaBase_220515_183509.pdf

Ciocca, D. R., & Delgado, G. (2017). The reality of scientific research in Latin America; an insider’s perspective. *Cell Stress and Chaperones*, 22(6), 847–852. <https://doi.org/10.1007/s12192-017-0815-8>

Cohen, B. (1963). *The Press and Foreign Policy*. Generic.

Colombia. (1991). Constitución política de Colombia. Alianza Editorial.

Colombia. (1994, July 11). Ley 142 de 1994. Estado Colombiano. Retrieved March 19, 2022, from <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=2752>

Colombia. (2012, October 17). Ley 1581 de 2012. Estado Colombiano. Retrieved March 22, 2022, from <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>

DANE. (2022, March 19). Departamento Administrativo Nacional de Estadísticas (DANE) [Press release].

https://www.dane.gov.co/files/geoestadistica/Preguntas_frecuentes_estratificacion.pdf

DANE. (2018, September). Censo Nacional de Población y Vivienda - CNPV 2018. <https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018>

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, 70(5), 979–995. <https://doi.org/10.1037/0022-3514.70.5.979>

Diario La República. (2017, May 27). Conozca cómo es el mapa de los estratos en las grandes ciudades de Colombia. Retrieved April 26, 2022, from <https://www.larepublica.co/economia/este-es-el-mapa-de-los-estratos-en-las-grandes-ciudades-del-pais-2866032>

Drentea, P. (2000). Age, Debt, and Anxiety. *Journal of Health and Social Behavior*, 41(4), 437. <https://doi.org/10.2307/2676296>

Duggan, M., & Smith, A. (2016). The political environment on social media. Pew Research Centre.

El Colombiano. (2022, May 10). "Golpe de Estado en Medellín": Petro cuestiona suspensión de Quintero. Retrieved May 19, 2022, from <https://www.elcolombiano.com/colombia/politica/gustavo-petro-dice-que-suspension-de-la-procuraduria-a-daniel-quintero-es-un-golpe-de-estado-en-medellin-if17475287>

El Espectador. (2022, May 12). ¿Se puede hablar de "golpe de Estado" tras la suspensión del alcalde de Medellín? Retrieved May 19, 2022, from <https://www.elespectador.com/politica/elecciones-colombia-2022/se-puede-hablar-de-golpe-de-estado-tras-la-suspension-del-alcalde-de-medellin/>

El Espectador. (2021, June 6). Las 180 leyes anticomunistas en Colombia que habrían azuzado el conflicto armado. El Espectador. Retrieved March 21, 2022, from <https://www.elespectador.com/colombia-20/paz-y-memoria/las-180-leyes-anticomunistas-en-colombia-que-habrian-azuzado-el-conflicto-armado-article/>

El Tiempo. (2022, April 2). Petro y Fico, adelante en la 1.a vuelta; para la 2.a se aprieta el duelo. Retrieved April 26, 2022, from <https://www.eltiempo.com/elecciones-2022/presidencia/encuesta-presidencial-petro-y-fico-sacan-distancia-a-los-otros-candidatos-662562>

Enns, P. K., Lagodny, J., & Schuldt, J. P. (2017). Understanding the 2016 US Presidential Polls: The Importance of Hidden Trump Supporters. *Statistics, Politics and Policy*, 8(1). <https://doi.org/10.1515/spp-2017-0003>

Eurostat. (2021, November 1). Quality of life indicators - education. Retrieved July 2, 2022, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Quality_of_life_indicators_-_education&oldid=548962#Education_in_the_context_of_quality_of_life

Frank, M. R., Cebrian, M., Pickard, G., & Rahwan, I. (2017). Validating Bayesian truth serum in large-scale online human experiments. *PloS one*, 12(5), e0177385.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35–42. <https://doi.org/10.1016/j.socec.2010.10.008>

Galesic, M., Bruine De Bruin, W., Dumas, M., Kapteyn, A., Darling, J. E., & Meijer, E. (2018). Asking about social circles improves election predictions. *Nature Human Behaviour*, 2(3), 187–193. <https://doi.org/10.1038/s41562-018-0302-y>

Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice. *Statistical Science*, 24(3). <https://doi.org/10.1214/09-sts301>

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Anchor.

Guarumo & Ecoanalitica. (2022, May 1). Petro y Fico, favoritos para primera vuelta, pero habría empate en segunda. *El Tiempo*. Retrieved May 18, 2022, from <https://www.eltiempo.com/elecciones-2022/presidencia/encuesta-de-guarumo-petro-y-fico-favoritos-para-primera-vuelta-668979>

Helzer, E. G., & Dunning, D. (2012). Why and when peer prediction is superior to self-prediction: The weight given to future aspiration versus past achievement. *Journal of Personality and Social Psychology*, 103(1), 38–53. <https://doi.org/10.1037/a0028124>

Hogg, M. (2001). *Blackwell Handbook of Social Psychology*. Chapter three: Social Categorization, Depersonalization, and Group Behavior (1st ed.). Wiley-Blackwell.

Internal Displacement Monitoring Centre (IDMC). (2020, October 31). The Last Refuge: Urban Displacement in Colombia. Retrieved March 19, 2022, from <https://story.internal-displacement.org/colombia-urban/index.html>

Invamer. (2022, April 29). Petro tocó techo y Gutiérrez es el único que crece: habrá segunda vuelta, encuesta Invamer. El Colombiano. Retrieved May 18, 2022, from <https://www.elcolombiano.com/colombia/politica/petro-y-gutierrez-se-medirian-en-segunda-vuelta-presidencial-encuesta-invamer-CB17343100>

INVAMER. (2018, June). Elecciones 2018 medición junio 8 2018. Revista Semana. <https://www.semana.com/nacion/articulo/encuesta-invamer-en-segunda-vuelta-petro-pierde-el-caribe-y-gana-en-bogota/570344/>

Irwin, G. A., & van Holsteyn, J. J. M. (2002). According to the Polls. *Public Opinion Quarterly*, 66(1), 92–104. <https://doi.org/10.1086/338632>

Israel, G. (1992). Determining Sample Size. University of Florida, 2–5.

Krauss, R. M. (1981). Impression formation, impression management, and nonverbal behaviors. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium* (Vol 1, pp. 323–341).

Krumpal, I. (2011). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47(4), 2025–2047. <https://doi.org/10.1007/s11135-011-9640-9>

Kuha, J. (2022, April 27). Why are polls important during elections? | LSE Research. LSE Research for the World. Retrieved July 1, 2022, from <https://www.lse.ac.uk/research/research-for-the-world/impact/the-politics-of-polling-why-are-polls-important-during-elections>

Kukla-Gryz, A., Tyrowicz, J., Krawczyk, M., & Siwiński, K. (2015). We all do it, but are we willing to admit? Incentivizing digital pirates' confessions. *Applied Economics Letters*, 22(3), 184- 188.

La Patria. (2022, June 10). La campaña presidencial es cada vez más sucia. Retrieved June 13, 2022, from <https://www.lapatria.com/nacional/la-campana-presidencial-es-cada-vez-mas-sucia-492137>

Latin American Public Opinion Project (LAPOP Lab) & Vanderbilt University. (2010). El variado significado de “izquierda” y “derecha” en América Latina. Vanderbilt University. Retrieved March 21, 2022, from <https://www.vanderbilt.edu/lapop/insights/I0838es.pdf>

Lazarsfeld, P. F., and R. K. Merton. (1954). "Friendship as a Social Process." Pp. 18-66 in *Freedom and Control in Modern Society*, edited by M. Berger, T. Abel, and C. H. Page. Princeton, N.J.: Van Nostrand.

Luhmann, N. (1996). *The Reality of the Mass Media*. Stanford University Press.

McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly*, 36(2), 176. <https://doi.org/10.1086/267990>

McQuail, D. (2010). *McQuail's Mass Communication Theory*. SAGE Publications Ltd.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>

Mendoza Torres, M. R., Tarazona Morales, O., & Duque Vargas, L. (2011). Caracterización de la pobreza oculta y su efecto en la sostenibilidad económica de las pequeñas y medianas empresas de la localidad de Usaquén (Bogotá - Colombia). *Revista Facultad de Ciencias Económicas*, 19(2), 123–140. <https://doi.org/10.18359/rfce.2252>

Mosteller, F., Hyman, H., McCarthy, P. J., Marks, E. S., & Truman, D. B. (1949). The Pre-Election Polls of 1948. Report to the Committee on Analysis of Pre-Election Polls and Forecasts. New York: Social Science Research Council., 23(3). <https://doi.org/10.2307/361429>

Nisbett, R. E., & Kunda, Z. (1985). Perception of social distributions. *Journal of Personality and Social Psychology*, 48(2), 297–311. <https://doi.org/10.1037/0022-3514.48.2.297>

Noelle-Neumann, E. (1974). The Spiral of Silence a Theory of Public Opinion. *Journal of Communication*, 24(2), 43–51. <https://doi.org/10.1111/j.1460-2466.1974.tb00367.x>

Oakes, J., & Rossi, P. H. (2003). The measurement of SES in health research: current practice and steps toward a new approach. *Social Science & Medicine*, 56(4), 769–784. [https://doi.org/10.1016/s0277-9536\(02\)00073-4](https://doi.org/10.1016/s0277-9536(02)00073-4)

Razón Pública. (2022, April 24). Saltar techos: Lo que dicen las encuestas en las elecciones presidenciales de 2022. Retrieved April 26, 2022, from <https://razonpublica.com/saltar-techos-lo-dicen-las-encuestas-las-elecciones-presidenciales-2022/>

Ritzer, G., Dean, P., & Jurgenson, N. (2012). The Coming of Age of the Prosumer. *American Behavioral Scientist*, 56(4), 379–398. <https://doi.org/10.1177/0002764211429368>

Registraduría Nacional del Estado Civil. (2022, May). Elecciones presidenciales 2022 (No. 22). <https://resultados.registraduria.gov.co/presidente/5/colombia/bogota-d.c.>

Registraduría Nacional del Estado Civil. (2022, April 29). 39 002 239 colombianos están habilitados para votar en las elecciones de presidente y vicepresidente de la república de 2022 Retrieved May 5, 2022, from <https://www.registraduria.gov.co/39-002-239-colombianos-estan-habilitados-para-votar-en-las-elecciones-de.html>

Smith, V. (1982). Microeconomic Systems as an Experimental Science. *The American Economic Review*, 72(5), 923–955.

Snibbe, A. C., & Markus, H. R. (2005). You Can't Always Get What You Want: Educational Attainment, Agency, and Choice. *Journal of Personality and Social Psychology*, 88(4), 703–720. <https://doi.org/10.1037/0022-3514.88.4.703>

Stokes, S. (2009). Globalization and the left in Latin America. Unpublished manuscript, Department of Political Science, Yale University.

Toffler, A. (1980). *The third wave*. New York, NY: William Morrow.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9)

Uakarn, C., Chaokromthong, K., & Sintao, N. (2021). Sample Size Estimation using Yamane and Cochran and Krejcie and Morgan and Green Formulas and Cohen Statistical Power Analysis by G*Power and Comparisons. *Aphait International Journal*, 10(2), 76–88. <https://so04.tci-thaijo.org/index.php/ATI/article/view/254253>

Valora Analitik. (2022, April 17). Fico amplía ventaja sobre Petro para elecciones presidenciales en Colombia, según PredictIt. Retrieved April 26, 2022, from <https://www.valoraanalitik.com/2022/04/17/fico-amplia-ventaja-sobre-petro-para-elecciones-presidenciales/>

Vargas, S. (2022, June 10). Así se ha comportado la campaña sucia contra Petro y Rodolfo en redes sociales. *W Radio*. Retrieved June 13, 2022, from <https://www.wradio.com.co/2022/06/10/asi-se-ha-comportado-la-campana-sucia-contra-petro-y-rodolfo-en-redes-sociales/>

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities* (2nd ed.). Wiley Interscience.

Weaver, R., & Prelec, D. (2013). Creating truth-telling incentives with the Bayesian truth serum. *Journal of Marketing Research*, 50(3), 289-302.

Williams, E. F., & Gilovich, T. (2008). Conceptions of the Self and Others Across Time. *Personality and Social Psychology Bulletin*, 34(8), 1037–1046. <https://doi.org/10.1177/0146167208317603>

World Bank. (2021, January 1). GDP per capita (current US\$) - Colombia | Data. World Bank Data. Retrieved March 17, 2022, from <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=CO>

Yamane, Taro. (1973), *Statistics: An Introductory Analysis*. London: John Weather Hill, Inc.

YanHaas. (2022, May 10). Resultados de la cuarta Gran Encuesta para las presidenciales. Noticias RCN. Retrieved May 18, 2022, from <https://www.noticiasrcn.com/elecciones/elecciones-2022-resultados-gran-encuesta-rcn-mayo-418885>

Appendix A

Survey

Consent

Muchas gracias por querer participar de la encuesta. **Esta hace parte de mi investigación para obtener el título de maestría.**

El objetivo es conocer su intención de voto en las próximas elecciones presidenciales que se llevarán acabo el 29 de mayo de 2022. **No hay respuestas correctas o incorrectas.** Las respuestas son completamente anónimas y toda información recolectada será tratada bajo la ley de **Habeas Data**.

En caso de tener cualquier duda sobre la encuesta me puede escribir al correo: 59720lnr@student.eur.nl

Nuevamente muchas gracias por ayudarme.

Nicolás Ramírez Saldarriaga

¿Acepta que es mayor de edad y los términos y condiciones de la encuesta descritos anteriormente?

Acepto

No Acepto

Thank you very much for answering this survey. It is part of my research for my Master's thesis. The goal is to get an idea of whom you are going to vote for in the next presidential elections that will happen Sunday the 22nd of May 2022. There are no right or wrong answers. All answers will be kept anonymously and any information collected will be under the Colombian law Habeas Data.

If you would like to contact me to discuss this research please email me at 597201nr@student.eur.nl.

Once again, thank you for helping me

Nicolás Ramírez Saldarriaga

Do you accept that you are at least 18 years old and the previously described terms and conditions?

I accept

I do not accept

Personal questions

¿Tiene planeado votar en las próximas elecciones presidenciales que se llevarán acabo el 29 de mayo de 2022?

Sí

No

Are you planning on voting in the upcoming election for president?

¿Por quién planea votar en las elecciones presidenciales de este 29 de mayo de 2022?

Gustavo Francisco Petro Urrego – Coalición Pacto Histórico

Sergio Fajardo Valderrama – Coalición Centro Esperanza

Federico Andrés Gutiérrez Zuluaga – Coalición Equipo por Colombia

Enrique Gómez Martínez – Partido Movimiento de Salvación Nacional

John Milton Rodríguez González – Partido Colombia Justa Libres

Rodolfo Hernández Suárez – Movimiento Liga de Gobernantes Anticorrupción

Voto en blanco

Ingrid Betancourt Pulido – Partido Verde Oxígeno

For whom are you planning to vote?

[This question was only visible if the respondent answered yes in the previous one]

Social circle questions

Piense en sus **amigos, familiares colegas y otros conocidos** que sean mayores de 18 años y con los que haya tenido algún tipo de **conversación en el último mes**, puede haber sido en persona o de cualquier otra forma. **Este grupo será su círculo social.**

¿Qué porcentaje de su círculo social cree que tiene planeado votar en las elecciones presidenciales que se llevarán a cabo el próximo 29 de mayo de 2022?

Por ejemplo, 0% significa que nadie dentro de su círculo social va a votar y 100% que todas las personas dentro de su círculo social van a votar.

En caso de no estar seguro, trate de dar un valor aproximado.

Deslice el cursor para indicar el porcentaje deseado

0 10 20 30 40 50 60 70 80 90 100



Think of your friends, family, colleagues, and other acquaintances 18 years of age or older that you have communicated with at least briefly within the last month, either face-to-face or otherwise. We will call these people your social contacts

What percentage of your social contacts is likely to vote in the upcoming election for President? For instance, 0% means that you think none of your social contacts will vote, and 100% means that all of your social contacts will vote. If you are not sure, just try to give your best guess.

Para la siguiente pregunta solo tenga en cuenta a aquellos dentro de su círculo social que planean votar en las elecciones presidenciales que se llevarán a cabo el próximo 29 de mayo de 2022

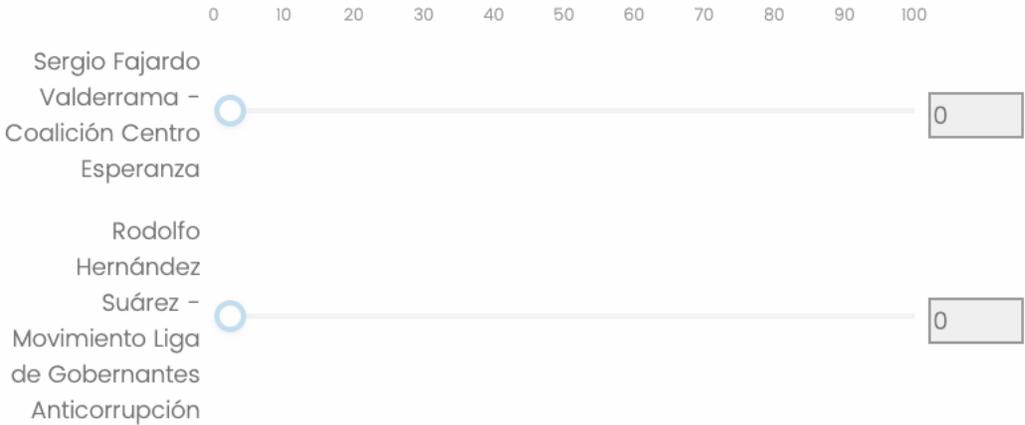
De su círculo social que planea votar ¿qué porcentaje cree que lo hará por los siguientes candidatos?

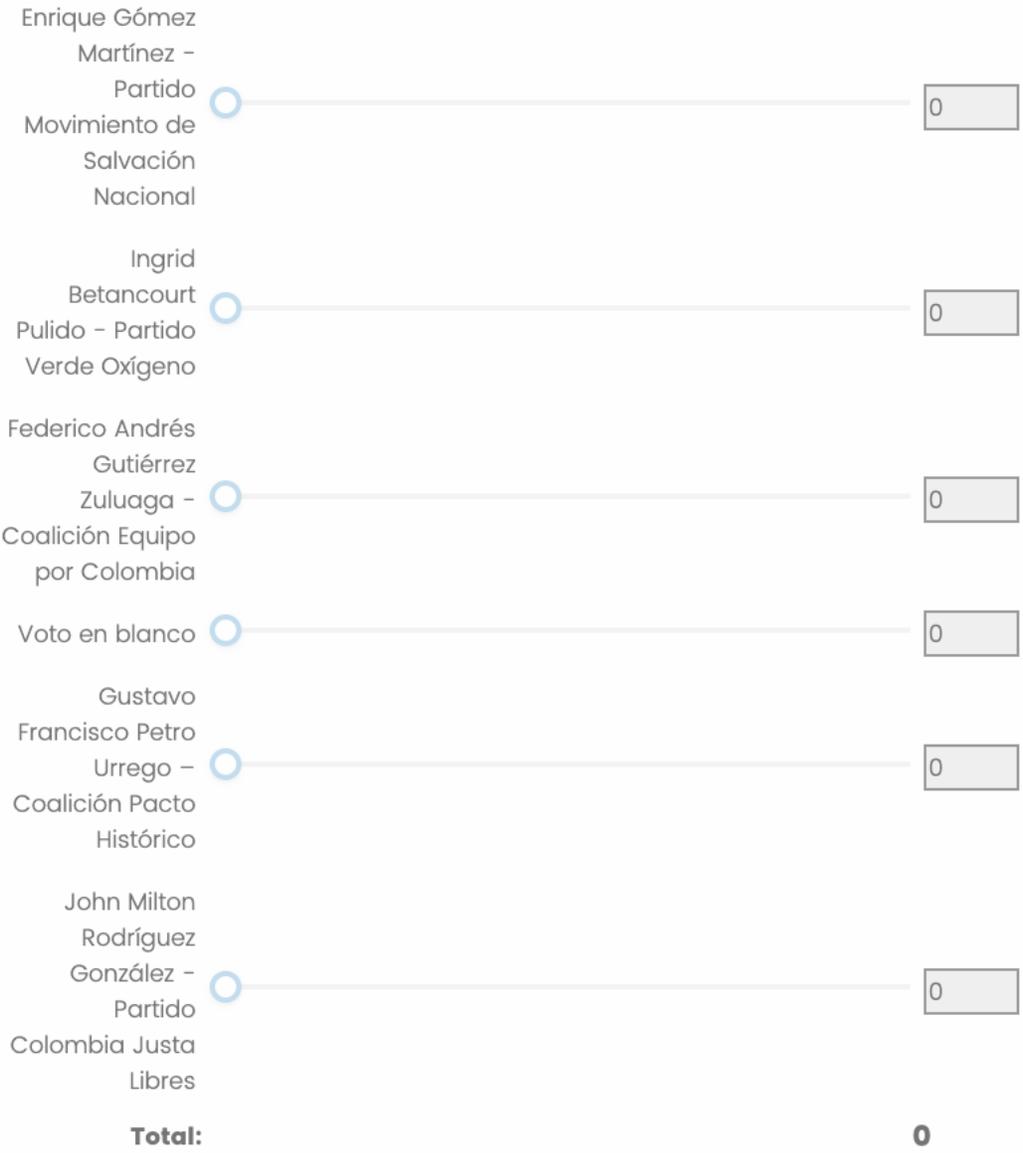
Por ejemplo, 0% significa que nadie dentro de su círculo social va a votar por ese candidato y 100% que todas las personas dentro de su círculo social van a votar por dicho candidato.

En caso de no estar seguro, trate de dar un valor aproximado.

La suma de todos los porcentajes seleccionados debe ser exactamente igual a 100%.

Deslice el cursor para seleccionar los porcentajes deseados





For the next question, please consider only those of your social contacts who are likely to vote in the upcoming elections. Of all your social contacts who are likely to vote, what percentage do you think will vote for the following candidates? For instance, 0% would mean that you think no voters in your social circle will vote for that candidate, and 100% means that all voters in your social circle will vote for that candidate. Again, if you are not sure, just try to give your best guess. The sum of all the percentages should be equal to 100%.

Demographics

Select the state where you live

[As Bogotá is the capital city is considered a state]

Selecione su departamento de residencia

Select your socioeconomic number

Selecione el estrato socioeconómico al que pertenece

1

2

3

4

5

6

Select your gender

Selecione su sexo

Masculino

Femenino

How old are you?

¿En qué grupo de edad se encuentra usted?

18 - 21

22 - 34

35 - 44

45 - 54

55 - 64

65 o más

What is your highest educational level achieved?

¿Cuál es el máximo nivel de estudios alcanzado?

Primaria

Bachillerato

Técnico

Tecnólogo

Pregrado

Maestría

Doctorado

Final question results

What do you think are going to be the results of the next presidential elections? Select the percentage that you think each candidate will obtain.

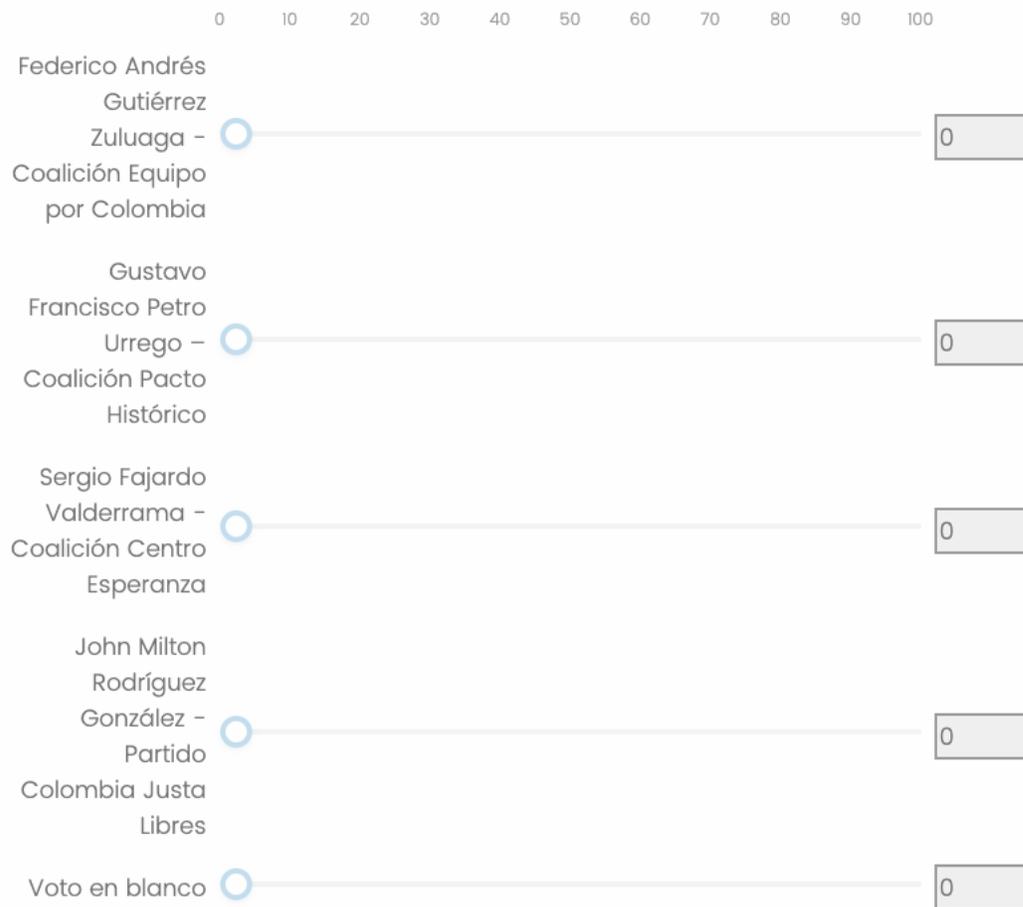
The sum of all the percentages should be equal to 100%.

¿Cuáles cree que serán los resultados de las elecciones el próximo domingo 29 de mayo de 2022?

Marque el resultado, en porcentaje, que cree que obtendrá cada uno de los candidatos

La suma de todos los porcentajes seleccionados debe exactamente igual a 100%.

Deslice el cursor para seleccionar los porcentajes deseados

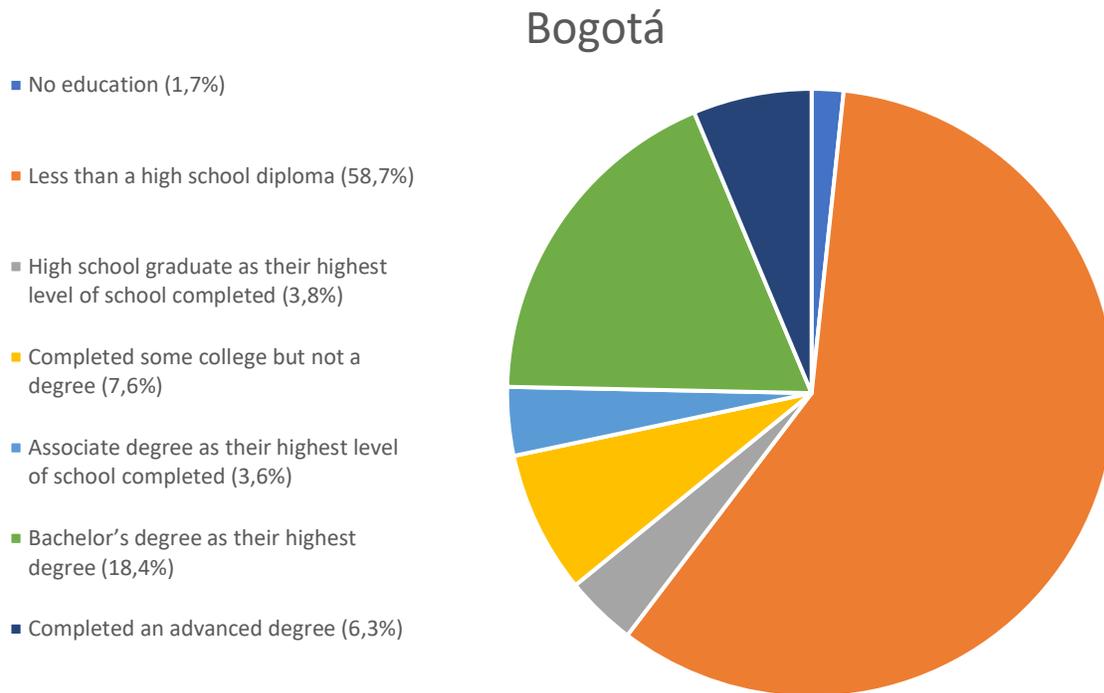


Enrique Gómez Martínez - Partido Movimiento de Salvación Nacional	<input type="text" value="0"/>
Rodolfo Hernández Suárez - Movimiento Liga de Gobernantes Anticorrupción	<input type="text" value="0"/>
<input type="text" value="0"/>	<input type="text" value="0"/>
Ingrid Betancourt Pulido - Partido Verde Oxígeno	<input type="text" value="0"/>
Total:	0

Appendix B

Education level Bogotá

According to DANE (2018), only 40% of Bogotá's population have a high school diploma or a higher degree.



Appendix C

Tables 7 to 10 show a description of the variables.

Table 7

Respondents' socioeconomic number distribution

<i>Socioeconomic Number</i>	<i># Answers</i>	<i>Percentage</i>
1	4	1,3%
2	46	14,4%
3	61	19,1%
4	75	23,5%
5	77	24,1%
6	56	17,6%
<i>Grand Total</i>	<i>319</i>	<i>100,0%</i>

Table 8

Respondents Gender distribution

<i>Gender</i>	<i># Answers</i>	<i>Percentage</i>
<i>Female</i>	<i>179</i>	<i>56,1%</i>
<i>Male</i>	<i>140</i>	<i>43,9%</i>
<i>Grand Total</i>	<i>319</i>	<i>100,0%</i>

Table 9

Respondents age distribution

<i>Age</i>	<i># Answers</i>	<i>Percentage</i>
<i>18 - 21</i>	<i>3</i>	<i>0,9%</i>
<i>22 - 34</i>	<i>41</i>	<i>12,9%</i>
<i>35 - 44</i>	<i>63</i>	<i>19,7%</i>
<i>45 - 54</i>	<i>78</i>	<i>24,5%</i>
<i>55 - 64</i>	<i>83</i>	<i>26%</i>
<i>65 or more</i>	<i>51</i>	<i>16%</i>

<i>Grand Total</i>	<i>319</i>	<i>100,0%</i>
--------------------	------------	---------------

Table 10

The highest educational level achieved respondents' distribution

<i>The highest educational level achieved</i>	<i># Answers</i>	<i>Percentage</i>
<i>High School</i>	<i>30</i>	<i>9,4%</i>
<i>Ph.D</i>	<i>11</i>	<i>3,4%</i>
<i>Master's</i>	<i>92</i>	<i>28,8%</i>
<i>Bachelor's</i>	<i>126</i>	<i>39,5%</i>
<i>Primary</i>	<i>7</i>	<i>2,2%</i>
<i>Technólogo (2-3 years of BA)</i>	<i>27</i>	<i>8,5%</i>
<i>Técnico (1 year of BA)</i>	<i>26</i>	<i>8,2%</i>
<i>Grand Total</i>	<i>319</i>	<i>100,0%</i>

Appendix D

Tables 11 to 12 show a description and comparison of the variables once the 88 responses were eliminated.

Table 11

Comparison of respondents' Socioeconomic number

<i>Socioeconomic Number</i>	<i># Answers</i>	<i>Percentage</i>	<i>Previous percentage</i>
1	3	1,31%	1,3%
2	39	17,03%	14,4%
3	43	18,78%	19,1%
4	53	23,14%	23,5%
5	57	24,89%	24,1%
6	34	14,85%	17,6%
Grand Total	229	100,00%	100,0%

Table 12

Comparison of respondents' gender distribution

<i>Gender</i>	<i># Answers</i>	<i>Percentage</i>	<i>Previous percentage</i>
Female	122	53,28%	56,1%
Male	107	46,72%	43,9%
Grand Total	229	100,00%	100,00%

Table 13

Comparison of respondents' age distribution

<i>Age</i>	<i># Answers</i>	<i>Percentage</i>	<i>Previous percentage</i>
18 - 21	2	0,87%	0,9%
22 - 34	36	15,72%	12,9%

35 - 44	45	19,65%	19,7%
45 - 54	60	26,20%	24,5%
55 - 64	52	22,71%	26%
65 or more	34	14,85%	16%
Grand Total	229	100,00%	100,0%

Table 14

Comparison of respondents' distribution of highest educational level achieved

The highest educational level achieved	# Answers	Percentage	<i>Previous percentage</i>
High School	27	11,79%	9,4%
Ph.D	7	3,06%	3,4%
Master's	69	30,13%	28,8%
Bachelor's	86	37,55%	39,5%
Primary	5	2,18%	2,2%
Technólogo (2-3 years of BA)	19	8,30%	8,5%
Técnico (1 year of BA)	16	6,99%	8,2%
Grand Total	229	100,00%	100,0%