Understanding Messages to Underrepresented Racial, Ethnic, Gender, and Sexual Groups on Social Media by Democratic Politicians and their Electoral Implications

Henry Smith

July 15, 2020

Summary of the Proposal

Social media has continued to proliferate, not only as a space for self-expression, but also for greater communication between politicians and constituents. In this project, we seek to answer the question of how Democratic politicians, who rely on a diverse base for support, effectively connect with and mobilize constituents belonging to underrepresented racial, gender and sexual groups through social media. To do so, we will collect a data set of the visual media posted by Democratic politicians and build a prediction model of variables measured from these images related to their diversity appeal as well as their more generalized traits. We will subsequently analyze how Democratic candidates in racially diverse districts present themselves through social media during contested primary elections.

Background

0.1 Introduction

The importance of underrepresented voters is not new to the Democratic party: a 2017 poll of registered voters by the Pew Research Institute of U.S. Politics and Policy estimated that only fifty-nine percent of self-identified Democrats/lean Democrats label themselves as white, compared to the eighty-nine percent of Republicans/lean Republicans. This figure is down from an estimated sixty-seven percent in 2007 and seventy-five percent in 1997. The same report approximates that Black voters constitute nineteen percent of this Democratic base, Hispanic voters twelve percent, and Asian together with other underrepresented racial/ethnic groups constitute ten percent [3].

Moreover, recent elections suggest the emergence of the LGBT community, which we classify as underrepresented gender and sexual individuals, as one of the most solid Democratic voting blocs. Exit polling by NBC following the 2018 midterm elections indicated that while LGBT voters constituted only six percent of the electorate, upwards of eighty-two percent of these voters supported the Democratic candidate [7].

Figure 1: A 2017 tweet from Massachusetts Senator Elizabeth Warren in support of the LGBT community at the Boston Pride Parade

Despite the distinct importance of these groups to the Democratic party, it is not clear that the party knows how to effectively mobilize underrepresented voters. This harrowing reality came to the forefront of the news cycle following a decade-low Black voter turnout during the 2016 general election [13]. In response to this fall in turnout, to which many have attributed Democratic presidential candidate Hillary Clinton's loss, the Democratic National Committee (DNC) pledged $2.5 million for the funding of programs to increase turnout among underrepresented groups during the 2018 midterm elections [12].

Of particular interest to our research is how Democratic politicians themselves aim to mobilize these communities through social media. Past research has underscored the importance of social media as spaces for underrepresented racial, gender, and sexual groups. In conflict with the narrative that a lack of access to technology divides disadvantaged racial groups, a recent

study has shown that online platforms in fact embolden social networks between these groups [10]. Likewise, it is estimated that eighty percent of LGBT adults engage on at least one social media website, which is much greater than the fifty-eight percent of the general public [1]. Acknowledging the heightened status of social media in the lives of these individuals, we see that our research will impart significant contributions to understanding voter and politician behavior.

0.2 Previous Works

Figure 2: A 2020 Facebook post by former Vice President and presidential hopeful Joe Biden regarding the importance of Black History Month

With social media emerging as a medium of popular political discourse among many of its users, numerous studies have detailed the ways in which politicians leverage these new channels of communication to their advantage. Due to the constant pressures of reelection, the politician must remain constantly perceptive to his/her constituency [6]. This idea, which forms the basis of American political theory, has been re imagined in the digital age. The attentiveness of the politician is no longer solely measured by mailers or presence at public events, but also his/her communication with constituents through social media platforms. Literature has reinforced this notion that social media has allowed for personalized dialogues between politicians and constituents, and that politicians are ultimately responsive to these conversations [5] [14].

Beyond the political capital of social media, we must additionally consider how politicians have appealed to underrepresented racial, sexual, and gender groups in electoral politics. Although numerous academics have tackled the polemic topic of voter identification laws and underrepresented voter turnout, few have appealed to the intraparty nature explaining why these groups are motivated to vote [11] [15]. Even so, research considering underrepresented voter turnout is often seemingly contradictory: while some claim that the candidate's race is not a deciding factor in turnout among underrepresented voters, others have tied diversity in political representation to greater underrepresented voter turnout [8] [16].

Goal and Objectives

Recognizing this gap in our understanding of intraparty appeal to underrepresented voters through social media and how these interactions drive these voters to the polls, we seek to answer the following questions:

1. How do Democratic politicians present themselves to underrepresented racial, gender, and sexual groups on social media platforms through visual content?
2. Which traits displayed in these images are perceived most positively/negatively by underrepresented voters?
3. How do visual messages predict primary election outcomes in racially diverse electoral districts?

In order to answer these questions, we will integrate knowledge of American electoral politics and machine learning techniques in the following ways:

- Perform large-scale collection of visual media from the social media accounts of Democratic politicians across one or more platforms (Instagram, Facebook, Twitter, etc.).
- Construct a training data set for these images using domestic crowd sourcing, consisting of both variables measuring appeal to underrepresented groups as well as general trait variables.
- Subsequently construct a neural network trained on the training image set for prediction modeling of these variables on the entire image data set. An accurate model will allow us to better understand the relationship between the variables we measured, providing particular insight into how Democratic politicians present themselves to underrepresented voters.

- Collect information on the electoral outcomes of contested Democratic primaries in racially diverse districts; in a similar manner, train a network to predict these outcomes based upon the set of variables for our images. This model, if accurate, will allow us to determine the success of particular methods employed by these politicians to mobilize underrepresented voters.

Methods

Many of the methods involved in such a study would involve the use of API's and neural networks to both collect and analyze large amounts of visual data. While I am admittedly not well-versed in either area, I have gained a basic understanding of the capabilities and applications of these technologies through my current coursework and by reading the most recent work ("Gender Stereotyped Visual Self-presentations of Politicians Can Predict Electoral Success") of my mentor, Dr. Jungseock Joo [4]. I conclude my proposal by presenting a basic outline of how I believe such a research process would take place:

1. **Image Collection and Interpretation**:

   –As in similar works published by the lab, the study would require the collection of images from at least one popular social media platform (Facebook, Twitter, Instagram). If possible, it would be interesting to compile data from across multiple social media sources to study if/how political messages vary across multiple platforms.

   –Having collected this information, the next task would be to classify the images by their appeal to specific underrepresented groups. This would most likely be done using a number of continuous variables scaled from 0 to 1, with an image rated '0' having no relation and '1' being most related to the respective group (including 'LGBT_appeal', 'Black_appeal', 'Hispanic_appeal', 'Asian_appeal', 'general_diversity_appeal', etc.). We would also want to account for the politician's known race/gender identity with the use of nominal variables. For instance, in Fig. 1 included above, the image would most likely receive a rating close to 1 on the 'LGBT_appeal' variable while closer to 0 on the 'Black_appeal' and 'Hispanic_appeal' variables. Provided this set of variables would ultimately be highly subjective, we would use a domestic crowd sourcing service to collect human responses on a subset of the images, our training set. This training set should be representative of the entire images data set so that, when performing predictions on images outside of the training set, our model is not extrapolating. Along with these particular "underrepresented appeal" variables, we would want to collect a set of more general political traits, such as 'qualified', 'competent', and 'elitist' (similar to those in "Gender Stereotyped Visual Self-presentations of Politicians Can Predict Electoral Success"), in order to assess the relationships between covariates in our model [4].

   –Subsequently, we would construct a network to predict this set of variables for the entire image data set. In order to do so, we would train the network on the set of crowd-sourced images and assess the accuracy of our model based upon this subset. The most interesting use of this model, however, would be the ability to perform multivariate statistical analyses on the data set of images to understand how Democratic politicians appeal to underrepresented voters, our original question posed. Options such as principal component analysis or cluster analysis may provide us with a better understanding of these conscious decisions Democratic politicians make when interacting with underrepresented groups. For instance, PCA may find that much of the variability in our original data is explained by the positive relationships between 'LGBT_appeal', 'friendly', and 'credible', indicating that Democratic politicians appeal to the LGBT community with messages that present them in a friendly and credible light. Moreover, techniques such as multivariate analysis of variance could impart insight into whether communication with underrepresented constituents varies depending upon a Democratic politician's own racial, gender, or sexual identity.

2. **Analysis of Primary Election Outcomes**:

   –Perhaps more interesting than how Democratic politicians communicate with underrepresented constituents is how these messages manifest during electoral competitions. We would like to study Democratic primary election outcomes as opposed to general election campaigns in order to gain a more complete picture of which strategies resonate with underrepresented voters.[1]

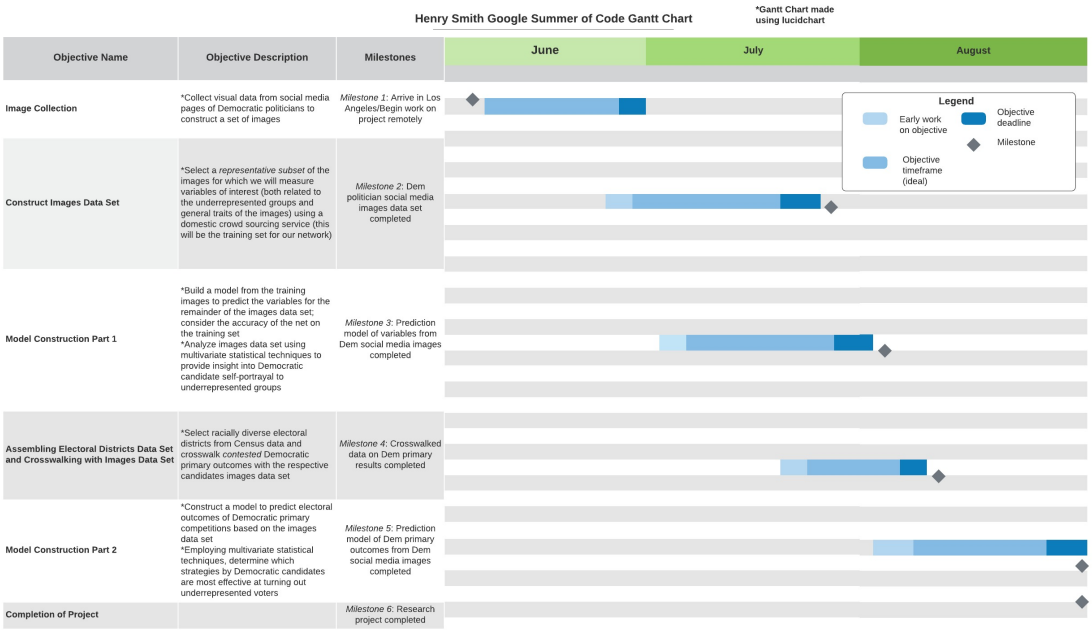
[1]Note: By looking at general elections, we would run into the presence of larger confounding variables, such as party identification, which would make it difficult to distinguish between the effect of the candidate's message and these other variables. However, primary elections also have historically low participation rates compared to general elections, meaning our results would be limited in scope: they may not be representative of the same effects of Democratic political messages to underrepresented voters during general elections.

–While we would obviously be unable to obtain individual primary election vote breakdowns by race, gender, and sexual identity, the United States Census Bureau does collect information about the racial makeup of individual congressional districts [2]. Accordingly, we could limit our focus to a number of contested primary elections in more racially diverse congressional districts. By selecting these districts, we would also be studying areas in which there is a greater importance for Democratic politicians to connect with underrepresented racial communities.

–Thereafter, we would crosswalk this data of election outcomes to our images data set. We accordingly subset the images data set twice: once to include only the candidates in districts in which we are interested, and again to only include images posted during during the election cycle. Lastly, we build a network to predict the difference in voting share between the two top primary finishers in each race from their social media presentations to underrepresented voters using the variables collected. Such an analysis would impart a better understanding of which candidate messages are preferred over others and to which messages underrepresented racial voters are most perceptive.

Tentative Timeline

Per the suggestion, I have included a Gantt chart illustrating the intended timeline of progress for the project throughout the summer [9]. This chart has also been attached along with my proposal in the case that it is difficult to read as part of the document.



As previously mentioned, I will be working with my mentor, Dr. Jungseock Joo, primarily contributing to the statistical analyses and political interpretations for the project, the areas in which I am most knowledgeable. Considering that my experience with using API's for collecting images and with prediction modeling using neural networks is not as extensive, this timeline is merely a visualization from my point of view. Some objectives may be more easy/difficult to achieve than anticipated depending upon the skills and expertise of team members.

References

[1] A survey of lgbt americans. *Pew Research Center*, 2013. Retrieved from: `https://www.pewsocialtrends.org/2013/06/13/a-survey-of-lgbt-americans/`.

[2] My congressional district 116th congress, 2018 american community survey 1-year estimates. *U.S. Census Bureau*, 2018. Retrieved from: `https://www.census.gov/mycd/`.

[3] Wide gender gap, growing educational divide in voters' party identification. *Pew Research Center*, 2018. Retrieved from: https://www.people-press.org/2018/03/20/wide-gender-gap-growing-educational-divide-in-voters-party-identification/.

[4] Anonymous Authors. Gender stereotyped visual self-presentations of politicians can predict electoral success. In progress.

[5] Pablo Barberá, Andreu Casa, Jonathan Nagler, Patrick J. Egan, Richard Bonneau, John T. Jost, and Joshua A. Tucker. Who leads? who follows? measure issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review*, 113(4):883–901, 2019.

[6] Robert J. Barro. The control of politicians: An economic model. *Public Choice*, 14:19–42, 1973.

[7] Tim Fitzsimons. Record lgbt support for democrats in midterms, nbc news exit poll shows. *NBC News*, 2018. Retrieved from: https://www.nbcnews.com/feature/nbc-out/record-lgbt-support-democrats-midterms-nbc-news-exit-poll-shows-n934211.

[8] Bernard L. Fraga. Candidates or districts? reevaluating the role of race in voter turnout. *American Journal of Political Science*, 60(1):97–122, 2016.

[9] HL Gantt. Work, wages and profit, published by the engineering magazine. *New York*, 1910.

[10] Amy L. Gonzales. Disadvantaged minorities' use of the internet to expand their social networks. *Communication Research*, 44(4):467–486, 2017.

[11] Zoltan Hajnal, Nazita Lajevardi, and Lindsay Nielson. Voter identification laws and the suppression of minority votes. *The Journal of Politics*, 79(2):363–379, 2017.

[12] Astead W. Herndon. Democrats plan new effort to target minority voters. *The New York Times*, 2018. Retrieved from: https://www.nytimes.com/2018/06/21/us/politics/democrats-minority-voters-midterms.html?smtyp=cur&smid=tw-nytimes.

[13] Jens M. Krogstad and Mark H. Lopez. Black voter turnout fell in 2016, even as a record number of americans cast ballots. *Pew Research Center*, 2017. Retrieved from: https://www.pewresearch.org/fact-tank/2017/05/12/black-voter-turnout-fell-in-2016-even-as-a-record-number-of-americans-cast-ballots/.

[14] Shannon C. McGregor. Personalization, social media, and voting: Effects of candidate self-personalization on vote intention. *New Media & Society*, 20(3):1139–1160, 2018.

[15] Jason D. Mycoff, Michael W. Wagner, and David C. Wilson. The empirical effects of voter-id laws: Present or absent? *Political Science & Politics*, 42(1):121–126, 2009.

[16] Rene R. Rocha, Caroline J. Tolbert, Daniel C. Bowen, and Christopher J. Clark. Race and turnout: Does descriptive representation in state legislatures increase minority voting? *Political Research Quarterly*, 63(4):890–907, 2010.

Henry Smith

Current Address: 23 Fourth St, Hanover, PA 17731

Phone Number: (717)-688-0103

Skype Email: henry.smith.skype@gmail.com

Email: henry.smith@yale.edu

EDUCATION

8.2018 - 5.2022 **Bachelor of Science in Statistics & Data Science** at Yale College, New Haven, CT

***Current GPA**: 3.95*
***Relevant Coursework**: Probability & Statistical Theory, Multivariate Statistics, Vector Analysis, Linear Algebra & Matrix Theory, Introduction to Computing & Programming, Game Theory & Political Science*
Expected graduation in May of 2022

8.2014 - 5.2018 **Diploma with Academic Honors** at Hanover Senior High School, Hanover, PA

***Cumulative GPA**: 5.04, Valedictorian*

WORK EXPERIENCE

fall 2019 **Learning Assistant for Introduction to Statistics: Political Science**

- Mentored a course of 50 undergraduate and graduate students in one-on-one and group office hour settings; developed interactive learning techniques to illustrate difficult course concepts; sought to create an inclusive and welcoming learning environment
- Graded weekly student homework assignments as well as course examinations; communicated with students to aid in understanding individual mistakes

2018 - present **Yale College Democrats**

- **Data Manager, Sergio Lopez for Campbell County, CA 2020 Campaign (fall 2019 - present)**– Oversaw data management for the campaign: used software to group upwards of 4,000 constituents into various target categories using demographic information; created voter lists that will be used in future campaign outreach
- **Yale Votes Constituent (fall 2018 - present)**– Assist in increasing student electoral awareness and participation on-campus by running voter registration and get out the vote events; talk with peers to dispel confusion surrounding the voter registration/absentee ballot request process
- **Legislative Advocacy Captain–Child Care (spring 2019)**– Conducted research on behalf of Connecticut House Bill 7014 and Raised Senate Bills 931, 933, and 934 to expand state childcare subsidies, which was presented to the chair of the Connecticut Appropriations Committee; spearheaded advocacy efforts by coordinating email-writing campaigns with on-campus student groups; garnered support among state senators and representatives on the Connecticut Joint Education Committee; authored an op-ed published in the *Connecticut Mirror* explaining and backing the passage of these bills
- **Data Team Fellow, Ned Lamont for Connecticut Governor Campaign (fall 2018)**– Volunteered at the New Haven campaign headquarters, utilizing DNC voter databases to determine high-interest individuals for get out the vote and volunteer recruitment efforts

2019 - present **Yale Volunteer Income Tax Assistance (VITA)**

- **Advanced Tax Preparer (fall 2019 - present)**– IRS-certified advanced tax preparer: complete state and federal tax returns for low-income New Haven residents with training in income (including business, capital, and retirement), deductions and credits, amended returns, and additional tax information; stimulate engagement with clients by walking through technical tax concepts step-by-step, remaining perceptive to concerns along the way
- **Intake Coordinator (spring 2019)**– Aided hundreds of clients with filing income tax returns at the New Haven Free Public Library; greeted and responded to questions from clients; managed copying and organization of income forms to pair clients with student tax assistants

SKILLS AND QUALIFICATIONS

Programming Languages

Intermediate skills	Python (Matplotlib, Numpy, Pandas, Pytorch, Pyspark), R, C, SQLite
Basic skills	CSS/HTML, JavaScript

Languages

Native	English
Advanced	Spanish

REFERENCES

Dr. Xiaofei (Susan) Wang (Professor of Intensive Introductory Statistics)

- wsusan@google.com

Dr. Christian Salas Pauliac (Professor of Game Theory & Political Science)

- chsalas@gmail.com