NATIONAL CHENG CHI UNIVERSITY

Big Data and Social Analysis

Course: 266868001

Analyzing Gender-Related Aspects of Donation Behavior for the US Election 2020

Prepared By:

Daniel Reichmann (110300604) [Writer]

Haruko Shimonishi (110204102) [Writer]

Lukas Groemer (110300603) [Writer]

Neo Shi Shiong (110703102) [Writer]

Sven Glinz (110308105) [Writer]

Prepared For: Professor Chung-pei Pien
16. June 2022

Contents

Introduction

The research question our group focuses on for the project is: "How does gender impact the election donation behavior across the US States?". The research gap is that US elections in 2020 are less investigated, especially on State level. According to Ondercin and Jones (2011) factors such as political knowledge can result in different levels of donation to political parties as compared to men. Further, Junn and Masuoka (2020) highlight that Gender plays a role in which party is predominately supported. Thus, we should be able to observe a difference in donation behavior between Female and Male donors in the United States. At the same time, analyzing Female Donation Behavior is becoming increasingly important since more and more focus is put on equal gender representation in a historically man dominated world. Together with a strongly increasing need for election donations, knowing the patterns and preferences of female donors could potentially be helpful in attracting them.

In addition, according to Open Secrets (2020) "Women broke barriers in the 2020 elections, both at the ballot box and as political donors". Compared to under 500Mio. in female donations to federal candidates in 2016, women almost twice as much to federal candidates in 2020, most of which went to Democrats. At the same time, the number of women running for congress more than doubled for both parties compared to 2016. Against this background, we analyze gender related donation patterns of the 2020 elections specifically for the two presidential Candidates Trump and Biden. We aggregate our numbers on national as well as state level, where especially the state level aggregation could provide new insights as major media outlets often only publish national numbers.

Data Sources & Cleaning

In order to analyze donation and gender related aspects of donations from private individuals to the Campaign Committees of Donald Trump and Joe Biden in 2020, we used the following Data:

Data	Source[1]	Description
Baby Names	SSA	Contains .txt files for every year (1880 - 2021) with 1.) **Name**, 2.) **Gender**, 3.) **Number of Births** of every born child in the United States.
Campaign Donations	FEC	Contains at least all contributions above 200$ to various US political committees and politicians. Includes among other **Full Name**, **Profession**, **Transaction Date** & **Amount**, **Address** & **Recipient of Donation.** In this paper, we only look at donations which went directly to the fundraising committees of Trump and Biden and are thus also subject to maximum donation amounts (currently XXX). Donations which went to other organizations which raised money for the two (various PACs etc.) are NOT considered.
Population Distribution by Gender	KFF	Contains the relative distribution of Male and Female in every US State for the year 2019.
US State Names and Abbreviations	50states.com	Contains the full Name as well as the Abbreviation for every US State and Territory
Political Ideology	Pew Research Center	Contains the % of people who identify as Conservative/ Moderate or Liberal for each US State in a survey from 2014.

Table 1: Description of Data Sources

As the total text file of donations is very large (> 16 GB) we first split it into smaller files as R must load each file into RAM. We read each one separately, only filter out the donations which went to the campaign committees of Biden and Trump and paste them together.

The most important part in order to present statistics on gender distribution of donations is to find a way how the gender for each donor can be identified, as the donation file only contains the field NAME which contains both first and last name. In order to identify the gender of each name we therefore do the following:

---

[1]For further Details, see appendix

> **Commented [Ga3]:** I think here we should explain a bit, how these ideologies relate to parties. Conservative is for me more or less republican, liberal is for me rather democrats. This is also reflected here similarly: https://news.gallup.com/poll/388988/political-ideology-steady-conservatives-moderates-tie.aspx

1. We first split "last name, first name" at the occurrence of the first comma. The first part becomes the last name, the second part the first name.

2. As we can only identify names which consist of one string without whitespace or non-alphanumeric characters, we clean the first name string by splitting it at every non-alphanumeric character and whitespace. Of all the received strings, **we select the longest one.**

NAME	LAST NAME	FIRST NAME	SPLIT BY NON-ALPHANUM.	LONGEST STRING	CLEANED FIRST NAME
HARLAN JR., ANDREW D.	HARLAN JR.	ANDREW D.	ANDREW / D	ANDREW	ANDREW
GUSTAFSON, JOHN & CONNIE	GUSTAFSON	JOHN & CONNIE	JOHN/ CONNIE	CONNIE	CONNIE
MEEK, CHARLES RONALD (RON)	MEEK	CHARKES RONALD (RON)	CHARLES / RONALD / RON	CHARLES	CHARLES

Table 2: Example for cleaning of Names

As we can see, sometimes names contain the name of the husband and spouse (JOHN & CONNIE). However, this is relatively rare and therefore it does not seem to pose a problem.

In a second step, we determine the gender of names with the files provided by the Social Security Administration. We take the birth years from 1940 to 2021, combine all files, group by names, and calculate for each name the percentage of cases in which it was used as a female or male name. If a name is used as Male (Female) in more than 75% of all cases, we determine it as Male (Female). All other names receive the gender NA as we cannot determine it with a high enough accuracy. To conclude, we join the Names and the associated gender onto the cleaned first names of the donations data frame and get the cleaned version of donations where each name is associated with a specific gender. We also remove all entries from the donation data set where the State abbreviation does not exist or is not part of the 50 US States + Territories (e.g. Some States such as AA (which refers to Military bases) are removed.) This only applies to a few entries. The abbreviations of the US States and Territories were scraped

from a table on the web. For our analysis, we also scrape information on the political ideology of each state and download a csv file which contains information on the gender distribution within US States.

Based on these files, all graphs can be produced by performing various joins, grouping by different variables and doing some calculations. For all graphs where we report relative donations of Male or Female, we use the total donations by Male and Female only (excluding donations where we could not identify the gender) as 100%.

Methods Used

1. Data Frame Manipulation

Data frame manipulation is one of the methods that is used throughout the project to sieve out the key information needed for our analysis from the huge amount of information in the donor's data. One of the ways it is done frequently is through the "*summarize*" and "*group by*" functions to collate the mass information into meaningful segregations.

2. Regression

Regression is another method that is important to our project as we attempt to find out if there are any relationships between the variables. In this project we aim to find out whether there is a correlation between the relative female donation amount depending on the conservativeness of a US State, on the percentage of its female population and most importantly depending on the candidate, Trump or Biden.

3. Time Series Analysis

Time series analysis is crucial when we want to see the change in the amount of donations by the donors for both candidates over the election period. We will be able to conduct more in-depth research through the donation pattern across time and attempt to find information to explain the particular donation pattern observed by the donors for the candidates.

5

4. Geographic Information System

It can be hard to make meaning out of the data that we got through our analysis, especially when the results involve the various states. GIS is one of the methods we adopt in the project in order to visualize the results in an understandable manner. An example will be showing the share of female donors for each candidate across the states. Having it on a heatmap will clearly display the percentage share along with the states involved when it comes to donations by female donors.

Results

The first analysis conducted was to compare the total and average donations as well as the total number of donors by gender for Trump and Biden. As we can see in Table 3, Biden received more donations in total, irrespective of the gender. However, comparing male and female doners, one can see that the discrepancy between the donation amounts is bigger when comparing male and females. Trump received 184 million USD from Men, or 83% of the 222 million USD which Biden received. On the other hand, Trump received 106 million USD from Women, or only 59% of the 179 million USD received by Biden. Biden's relative share of female donations is therefore higher than Trumps, suggesting that Biden is more accepted by the female population. At the same time, we can also see that Men make bigger donations on average than Women for both Trump and Biden.

We can further see that we were unable to identify with high enough accuracy the gender of around 70'000 donors which donated a total of around 31 million USD. This amounts to a margin of uncertainty of 4%.

Gender	Candidate	Avg. Donation	Total Donations (Mio)	Total Donors (T)
F	Biden	451	179	397
F	Trump	313	106	340
M	Biden	587	222	377
M	Trump	385	184	477
NA	Biden	483	16	33
NA	Trump	400	15	38

Table 3: Summary Statistics of analyzed Data

Figure 4 again shows the relative amount of female donations and donors to both candidates as a total, this time excluding the values where we could not identify the gender. We can see that for both candidates, the percentage of female donors is higher than the total amount donated,
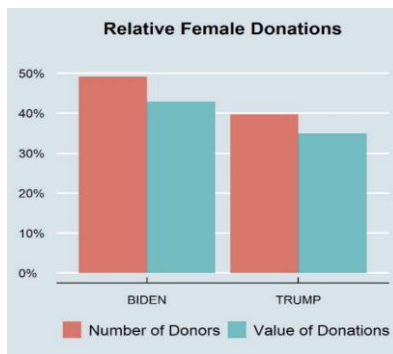

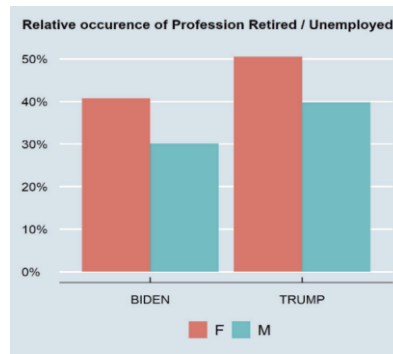*Figure 5: Relative Female Donations (Nation-wide), own depiction*

Figure 4: Relative Amount of Donations by Retirees (Nation-wide), own depiction

confirming the findings from the table above. Biden has relatively more female donors and received relatively more donations from female donors.

Looking at the profession of the donors in Figure 5, one can also gain a valuable insight: We see that a remarkable share of donations is made by retired people. First, we see that among the female donors there seem to be relatively more retirees than among the male donors (roughly 10% difference for both candidates). Further, we also see that among Trump's donors, a much higher share was retired. This is well in line with many media reports which

claim that Trump and the Republican Party enjoyed high popularity among the elderly. (npr.org, 2020).

Looking at the distribution of the relative donation share of females[2] in different US States on in Figure 6, we can see that Biden has almost no hotspots, except in the middle of the US. Trump is more dominant with females in the south of the USA. However, Biden's range of
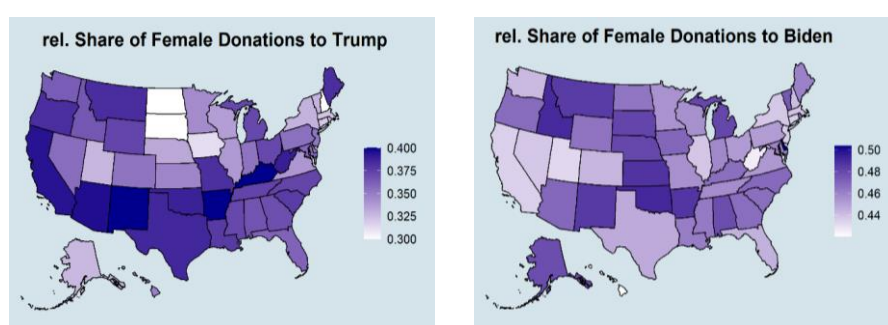

Figure 6: Relative Female Donations (State-level), own depiction

relative female donations is between 42 and 48 percent and has therefore a fluctuation of 6 percent, whereas trumps range is over 10 percent, from less than 30 percent to around 40 percent. In this context, it is important to highlight that the two maps use different scales and that in fact, Biden received relatively more donations from Females in **every** US State compared to Trump (as will become visible better in Figure 8 below). Thus, it is important to only compare the color coding within the maps and not across both maps[3]

In Figure 6, we can also see that the relative share of female donations by state varies relatively strongly between different states for both candidates. We are thus also interested in what could cause these differences. First, we believe that the percentage of female donors could be influenced by the percentage of female inhabitants per state[4]. Further, we believe that the

Commented [Ga4]: Writing % as percentage might be longer and more "formal"?

Commented [SG5R4]: percentage

---

[2] Rel. Share of Female Donations per Candidate calculated as: *Total Female Donations to Candidate X / Total Donations to Candidate X*

[3] E.g. just because California is darker in Trumps Map does not mean that he received a higher percentage of donations from women than Biden!

[4] which ranges from a max. of 52,5% Female in Washington DC to a min. of 49,1% Female in Wyoming

percentage of female donors could be influenced by the conservativeness of the state, where we believe that in more liberal states, relatively more women would donate and in less liberal, more conservative states, relatively less women would donate for a candidate.

We therefore set up the following models:

$$\%Donors\ F_i = \beta_0 + \alpha_0 * Candidate^5 + \beta_n * v + \epsilon_i$$

$$\%Donations\ F_i = \beta_0 + \alpha_0 * Candidate + \beta_n * v + \epsilon_i$$

Where our control variables v includes the Female Population Ratio per State (Regression 1 and 3 in Figure 7) and the Female Population Ratio per State as well as the percentage of people who classify as Conservative per State (Regression 2 and 4 in Figure 7).

	Dependent variable:			
	FEMALE RATIO			
	Number of Donors		Donation Amount	
	(1)	(2)	(3)	(4)
Female Ratio Population	0.880***	1.040***	0.486	0.638*
	(0.304)	(0.294)	(0.350)	(0.344)
Candidate: Trump	-0.115***	-0.115***	-0.106***	-0.106***
	(0.005)	(0.005)	(0.005)	(0.005)
Conservative		0.106***		0.102***
		(0.032)		(0.037)
Constant	0.072	-0.049	0.212	0.097
	(0.155)	(0.152)	(0.178)	(0.178)
Observations	102	102	102	102
R^2	0.857	0.872	0.791	0.805
Adjusted R^2	0.854	0.868	0.787	0.799
Residual Std. Error	0.024 (df = 99)	0.023 (df = 98)	0.028 (df = 99)	0.027 (df = 98)
F Statistic	296.947*** (df = 2; 99)	221.820*** (df = 3; 98)	187.057*** (df = 2; 99)	135.208*** (df = 3; 98)
Note:				$^*p<0.1;\ ^{**}p<0.05;\ ^{***}p<0.01$

Figure 7: Female Donations vs. Conservativeness: Regression

First, we observe what we have already shown in Figure 4 on national level, namely that the female contribution ratio is significantly higher for Biden than for Trump. (Both measured by

---

[5] Dummy Variable (Trump = 1, Biden = 0)

donation amount and number of donors). This is not only true on average on a national level but also holds for every single US state as can be seen very well in graph Figure 8 with a statistically significant difference of around 10% points for both measures (or around 0.2% / % Female Inhabitants as depicted in Figure 4) as revealed by the regression analysis.

Pop. Adj. %Female Donors vs. Conservativeness
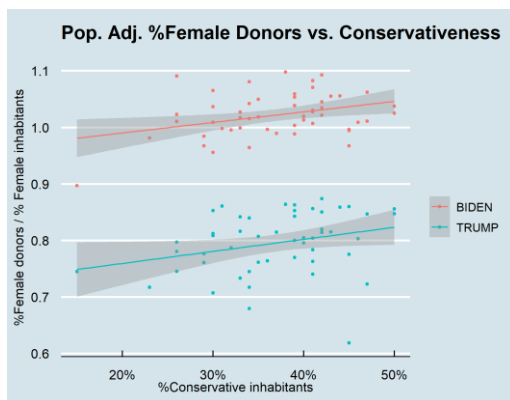
Further we observe that the female ratio in a state is statistically significant on the 1% level and strongly correlates with the % of female donors. On average, a 1% higher female ratio in a State means a 0.88 - 1% higher female donor ratio depending on the model specification. Interestingly, this relationship cannot be observed when taking the percentage of female donations (regressions 3-4, Figure 7) by value as the dependent variable, which intuitively makes sense since the donation size this is probably less dependent on the relative share of inhabitants but more on other factors (such as Income).

Ultimately, as we can see in regressions 2 and 4, both the percentage share of female donations and female donors correlate significantly positively with the percentage of people who identify themselves as conservative. This disproves our assumption that relatively more women would donate in liberal states, as the female donation ratio in fact increases the more people identify as conservative in a state.

Commented [Ga7]: Again Figure 7?

Commented [Ga8]: Correlates significantly positively or negatively? It may be in the figure explained, but maybe this is also important here.

Commented [SG9R8]: positively

Of course, it would be rash to determine conservativeness as a true determinant of female donor share since we could well have left out relevant control variables which lead to biased results. For example, there could be many reasons why the ratio of female donors and donations is higher in conservative states and surely, conservativeness by itself is not at the root of this. However, it is indeed interesting to see that the ratio of female donations and donors in conservative states is higher and this may warrant further investigation of what is responsible for this observation.
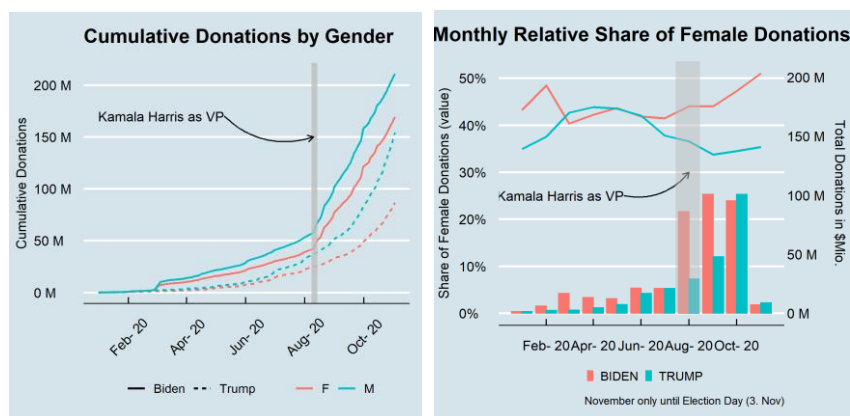

Figure 10: Cumulative Donations by Gender (own depiction)   Figure 9: Monthly relative Share of Female Donations (own depiction)

When we look at donations over time in a time-series analysis, we can see a sharp increase in total donations received by Biden and an increased female donation share beginning in the Month of August (from an initial 40% to over 50% by Election Day!). This could be argued by the naming of Kamala Harris as Vice President. This finding is also supported by the US news (US News, 2020). At the same time, the relative donation share by females contributing to Trumps campaign decreased. This could indeed mean that the naming of Kamala Harris as VP appealed to the female US citizens. However, as we can see in Graphs 9 and 10, it is also important to mention that absolute donations from Men also rose strongly after the naming Harris as the pick for Vice President and not only those from Women.

Conclusion

For examining the research gap, various data sources have been evaluated, transformed and integrated in the right formats to analyze the issues in question. For example, the assignment of genders to the donors was a challenge which was overcome by the support of an additional data source. The usage of these diverse data sources seems to have returned valuable insights by analyzing it on correlations, variations along the time series or location-based differences. This report can contribute to existing literature by presenting that woman were more likely to donate for the democratic candidate Biden in absolute as well as relative terms . However, the analysis also showed that women are donating less on average than men. The reasons behind this are still subject to further research and could possibly involve the professions or the average wealth of women versus men. Against our assumptions, that share of female donors per Candidate are increasing with the degree of liberalism within a state - our analysis suggests that female donors' share is increasing with the conservativeness of a US State. . Future studies can start from here and try to verify this outcome as well as search for the reason why relatively more women donate in conservative states. Interestingly, we also see that a larger share of Trump's donors were retired, whereby more female than male donors were retirees for both candidates. It could thus also be a point of further research to analyze why less non-retired women donate than men.

Furthermore, the time-series analysis suggests that the democrats landed a coup with the proposed appointment of Kamal Harris as VP. This is because female donor behavior evolved very positively for Biden following the An interesting field for further research, especially in the US, would be a race-based analysis in this context. This is because Kamal Harris is, as we know by now, the first Black and South Asian American elected as the VP of the US (The White House, 2022). On top of that, for further research, a data source containing the donors including the genders would increase the number of donations to analysis and thereby the

accuracy of analysis. The contemporary issue of a non-binary gender system could also be addressed in this context.

Overall, the report presents interesting implications for further research. Especially, as the relevance of this topic is not expected to decrease with election donations and spendings above one billion dollars (Allison et al., 2016). Further research could also include the analysis of gender trends in previous elections. Further as mentioned above in Table 1, we only analyzed the donations which went directly to the candidates' election campaign and omitted the vast amount of donations indirectly flowing to Candidates via Party Committees or Political Action Committees (PACs). Thus, future research could also include these donations and see if that leads to different results.

Commented [SG11]: I do not think that this exists anywhere and also cannot be generated bc. Donors do not have to reveal their gender on the donation form -> Should we take it out ?

References

50states.com. (2022, 06 01). Retrieved from https://www.50states.com/abbreviations.htm

Allison, B., Rojanasakul, M., Brittany, H., & Cedric, S. (2016). *Tracking the 2016 Presidential Money Race*. Retrieved 06 01, 2022, from Bloomberg.com: http://www.bloomberg.com/politics/graphics/2016-presidential-campaign-fundraising/

Federal Election Commission. (2022). *Bulk Data Download*. Retrieved 06 01, 2022, from https://www.fec.gov/data/browse-data/?tab=bulk-data

Junn, J., & Masuoka, N. (2020). The Gender Gap Is a Race Gap: Women Coters in US Presidential Elections. *Perspectives on Politics, 18*(4), 1135-45.

KFF. (2019). *Population Distribution by Sex*. Retrieved 06 01, 2022, from https://www.kff.org/other/state-indicator/distribution-by-sex/?currentTimeframe=0&sortModel={"colId":"Location","sort":"asc"}

npr.org. (2020). *Older White Voters Who Helped Trump Win May Be In Play*. Retrieved from https://www.npr.org/2020/08/23/904755484/older-white-voters-who-helped-trump-win-may-be-in-play?t=1655278417045

Ondercin, H. L., & White, D. (2011). Gender Jeopardy: What is the Impact of Gender Differences in Political Knowledge on Political Participation? *Social Science Quartlerly, 92*(3), 675-94.

Open Secrets. (2020). *In 2020 women ran, won and donated in record numbers*. Retrieved from https://www.opensecrets.org/news/2020/12/women-ran-won-donate-record-numbers-2020-nimp/

OUR GROUP (Name???). (2022). *GitHub Repository - Source Code*. Retrieved from https://github.com/svensglinz/Data_Analysis_Project

Pew Research Center. (2022). *Political ideology by state*. Retrieved 06 01, 2022, from https://www.pewresearch.org/religion/religious-landscape-study/compare/political-ideology/by/state/

Social Security Agency. (2022). *Pupular Baby Names - Beyond the Top 1000 Names*. Retrieved 06 01, 2022, from https://www.ssa.gov/oact/babynames/limits.html

The White House. (2022). *Kamala Harris - The Vice President*. Retrieved from https://www.whitehouse.gov/administration/vice-president-harris/

US News. (2020). *Bidne Sees Major Fundraising Surge Since Naming Kamala Harris as VP*.
Retrieved 06 01, 2022, from https://www.usnews.com/news/elections/articles/2020-
08-14/biden-sees-major-fundraising-surge-since-naming-kamala-harris-as-vp-pick

Code

All the code and data to reproduce the analysis of this paper can be found in our GitHub
Repository

https://github.com/svensglinz/Data_Analysis_Project