

Women Voters in Pakistani Politics

Abstract— Pakistan a Muslim country located in southern Asia. This research focuses on the impact of that women have had on party dichotomies during recent National Assembly elections. The paper suggests the existence of a multi-party that has seen an increasing ratio of women voters that may be increasing the popularity of Pakistan parties such as the PTI and PPP. We also suggest that recent decline within independent and small party movements in Pakistan is associated with a corresponding increase in ratio of votes coming from women. The paper has also discovered various factors that been associated with high women voting ratios such as high population, recency of election and isolation from the FATA province.

I. INTRODUCTION

A. Research Overview

In order to better understand foreign relations and global politics it is necessary to understand the internal affairs of the various countries that make up the global political system. One such country who is commonly known however not rarely known about internally is Pakistan which this research is focused on. In particular the scope of this research has been narrowed down to the potential influence that women voters have played on Pakistani politics. In hope that doing so can provide an understanding changing dichotomies of the country and give insight on events within Pakistan. The key questions of this research include: Who are the major Pakistani Parties? How have these party dichotomies changed over time? What factors are correlated with a high ratio of voters being women? How might the women voters be influencing Pakistani party dichotomies?

B. Background Knowledge

Pakistan is a Muslim nation located in southern Asia in between India, Afghanistan, and Iran. The country was formed in 1947 from the partitioning of the British India between it and its Hindu neighbor India. The nation's government is a federal parliamentary republic and is divided into 4 separate provinces along with a capital region and 2 Pakistani administered areas.¹

The country is split into 3 branches of government an executive, legislative, and judicial branch. First is the executive branch which contains a President who serves as head of state and is elected via an electoral college consisting of members of the legislative body. The executive branch also contains prime minister who serves alongside the president as the head of government.²

The legislative branch is bicameral in nature containing an upper house called the Senate which distributes its representation among the 4 provinces alongside seats given to capital region and the Federally Administered Tribal Areas.²

Pakistan also has a lower chamber called the National Assembly by Pakistani constitution members of the National Assembly are democratically elected through universal

suffrage where all above the age of 21 can vote. Members of the National Assembly serve in terms of 5 years as long as no external factors interrupt. 5% of National Assembly seats are reserved for specific minority groups in Pakistan who's elections run alongside the major seats at separate electorates. Seats for the National Assembly are distributed on a province basis. The Prime Minister and Cabinet positions are chosen from the members of the National Assembly by the president. This research will be focused on data oriented around the National Assembly due to the its historically democratic nature and availability of data.²

C. Political Parties

During this research 3 major Pakistani political parties will be explored. The Pakistani People's Party (PPP), The Pakistan Muslim League-Naswaz (PML-N) and The Pakistan Tehreek-e-Insaf (PTI).

Firstly, The Pakistani People's Party otherwise known as PPP is a left wing populist party that emerged during the late 60s and early 70s. The party has historically been deeply tied to the charismatic and popular Bhutto family having both Presidents and Prime Ministers including Pakistan's first woman Prime Minister in 1988. The parties ideology was initially built around anti-establishment rhetoric, Marxist, and anti-West. These ideals and there influence originated from a grassroots disturbed by Pakistan's failure in various foreign conflicts. In recent times however, the PPP has moderated much of its image in areas such as economics.³

Second, The Pakistan Muslim League-Naswaz otherwise known as the PML-N is regarded as a right wing part. Was formed as a offshoot of sorts of the old Muslim League by Naswaz Sharif. The party was formed as a large scale coalition of various factions opposed to the PPP such as religious groups, Businesses, and the Military. The party has supported more pro-West, free market, and Islamic religious values. The party has since had declining relations between it and the military in recent elections.³

Lastly, The Pakistan Tehreek-e-Insaf otherwise known as the PTI is a more moderate anti-establishment party. The most recently created of these parties the PTI was formed in 1996 by Pakistani cricketer and celebrity Ihmran Khan. The party prides itself on being a party of outsiders rather than politicians and criticizes the PPP and PML-N as being corrupt. they argue for more transparent and meritocratic systems of selecting government officials and operate overall as a catch-all party with a wider range of appeal. The party has found success among various businesses and business leaders as well as students and various political figures they have been taking in.³

D. Tools & Dataset

The tool behind this research is the extremely popular programming language Python. It is a practical for statistical research such as this due to its simplistic syntax, massive array

of various packages and libraries that can remove the need for heavy statistical programs to be developed. The language allows for seamless usage with Jupyter Notebook which allows for strong code partitioning and running code section at a time which is practical for the modular yet interconnected research done here.

The packages used here are Pandas and Geopandas which allow for easy usage with the datasets given both geographical and statistical that we are given access too. Matplotlib and Seaborn were used in order to generate the visuals to display various trends and images shown throughout the paper. Scipy is used in order to do many of the more complex hypothesis testing used to tests observations coming from the graphs. Shapely was used to convert various geological data into practical numerical data to extend the scope of the given datasets.

The dataset used on this obtained on Kaggle however data originated from Dawn a Pakistani newspaper that is written in english. The dataset is contains information about specific National Assembly seats and the candidate who won it. It has the following features: province, district, population, number of polling stations, number of male voters for seat, number of female voters for seat, total number of voters for seat, name of seat, number of votes obtained by winning candidate, candidates name, party affiliation, party, election year, Geojson data relating to seat.⁴

The rest of this paper is organized as follows. Section II various statistical and computational techniques used to generate images and acquire data. Section III displays and explains the resulting knowledge gained from the research. Section IV concludes the research

II. METHODS

A. Preprocessing

The first task to allow data to be applied in required visualizations and models created numerical data must be converted from its string based word format and instead stored as its corresponding integer. The numerical data categories are all hard coded into the pandas dataframe and all values in those categories are changed to there numerical counterparts.

Due to the nature of this research it is critical to have as comprehensive of a dataset as possible specifically relating the proper amount of Pakistani voters who are women. Therefore the first task executed by code takes the assumption that $\text{number of female voters} + \text{number of male voters} = \text{number of total voters}$ and since our dataset contains all 3 features any missing data points can be replaced via some arrangement of that formula. Therefore my program goes through each column and any missing columns identifies which data points are missing what data is had and calculates the proper data-point correspondingly.

It is also desired to be able to work across as many features as possible and thus generation of new features from what was originally provided came across as a beneficial and practical

idea. Two main approaches were used to generate these features firstly was manipulating the coordinates given via the geo-panda column and calculating relevant data from it second is taking the existing data points and extrapolating out new features that are combinations of various existing ones. For the first approach the algorithms given by the Shapely library expect there data in EPSG:3857 coordinate system otherwise known as the “Web Mercator” system which uses a flat plane representation of earth and uses meters as its unit of measurement while provided dataset is based on the EPSG:4326 “WGS84” coordinate system which is mapped onto a spherical representation of earth and uses degrees as the unit of measurement. Using a separate library PyProj which is designed to project coordinate systems into other coordinate systems this conversion is simplistic and easy. Upon this Shapely was able to calculate various piece of information about provinces such as there area, central x coordinate, and central y coordinate.

Using the other features in the dataset I also calculated the percentage of voters who are women which is one of the most central components to research. The Z-score of that ratio and a true or false boolean category that holds whether to consider that data-point an outlier where an outlier is any point greater than 2 standard deviations away from mean.

It is also noticeable with the initial dataset that two major factors relating to political parties that made for an not optimal usage display of data that being the existence of a large sum a parties with very few seats and a split in PTI between candidates within the party and candidates simply backed. I have found it logical for the latter to consider all candidates PTI-backed or part of party to be considered simply as a member of PTI for simplicity. Then for first problem only the top 6 political parties while all parties smaller along with independents were merged into a singular other category allowing for easier displays and removing a lot of noise from the major parties being analyzed. Lastly, for some statistics when parties are being compared to the national average an extra national “party” which consists of all seats was created and displayed.

B. Generating Baseline Understanding

Before looking the affect of women on politics it is necessary to gain a baseline understanding of the situation at which Pakistan is currently in and how its been changing politically. The first two question of this research are oriented towards this idea that being: Who are the major Pakistani Parties? How have these party dichotomies changed over time?

For the first question I have opted to represent the parties with a pie chart the idea here is that this sort of diagram is designed at displaying the distribution and ratio of various things which fits well for this question as it will easily show which Pakistani parties make up the largest piece of the

combined Pakistan electorate. Accomplishing this task is trivial by simply using the built in number of occurrences function to generate amount of times a specific party appears and feeding that data into Matplotlib's pie function.

In order to get an idea of the how the political parties have evolved and giving insight into the second question the most obvious place to start is what parties have been growing in recent times. This question we aimed at answering with a line graph to easily show the direction curve where the x axis represents the election year and y axis is number of seats that a party has won. The approach of this one changed from the last one due to the need to group together the data-points and then getting how many times that party shows up as the winning party. To do this I used the group function on both election year and party affiliation and the counting how many data-points are in each group which corresponds to how many seats were won by that party in that year followed by this data going into the line graph function by Matplotlib which generates the final visual.

The last of the visualizations that seemed necessary before entering into the women oriented features is the regional dominance of each party and where they are located this was done with two way firstly due to the Geo-JSON data available map generation was an available option and party could simply be done by color. However while visually appealing compounding with more statistical tools to confirm the legitimacy of the claims generated from it I created a similar map that used the center x and y and plotted a singular point onto a scatter plot to also represent the locational influence of a party. This is beneficial as by representing parties by 2 numbers and the categorical data point means we can run regression models on the data. In this case I have gone for SVC due to its simplicity and speed especially since accuracy is not too important given this is more for the baseline idea of location rather than a true prediction model. Finally to show these results to reader adding a contour to the data points makes the geographical divides quite clear.

C. Constructing Women Graphs

The initial starting point for analyzing women ratios in Pakistan was determining the basic statistic such as mean, median, the distribution curve, and whether the data is skewed. For the first 2 stats both are provided by a simple to use python library known as statistics. These two stats can also be used to determine skewed-ness of the distribution. This only leaves only testing for normality to which there are 3 methods implemented for firstly is simply drawing the histogram of the women ratio column. Second is the QQ-plot which allows for more accurate assessment of normality and lastly we can use the Shapiro-Wilkes Test which will give a p-value for if its normal or not. All three were implemented the first 2 due to there visual nature and allowing a better understanding to the reader and the latter to confirm the visuals line up with the reality of the curves shape.

Next in attempt to start working on the last question of how might women voters be affecting the Pakistani elections it would be optimal to be able to see which parties women are voting for however that data isn't available and therefore we have chosen to explore the association between a seats woman ratio and the party that had won there. Displaying this came in two graphs both of which were swarm plots with box plots overlaid on top of this. This was done so that mean women ratio of each party could be clearly identified alongside a background visual to give context to the box plot. This graph has been separated into 2 separate ones so one graph can show a closeup of mean to easily notice the difference in mean and a separate full view plot which allows user to see all points on the graph. Alongside this we have run hypothesis testing using the Mann-Whitney test to determine the likelihood that the data is above or below national mean on average. This test was chosen based on the corresponding distribution curve and data available to us.

Next going back to the third question at hand is: What factors are correlated with a high ratio of voters being women? For this I ran a correlation matrix between women ratio and various features using dummy variables where applicable. Several features however were not considered such as party due to the previous comparisons already made factors such as seat name which are individual IDs and aren't relevant to national correlation and factors who's correlations with women ratios are obvious such as ratio z-score and number of women votes. With the correlation matrix I selected the female ratio column only and used the Seaborn library to generate visually appealing graphic showing the correlations between the various features and women ratio.

This data was followed with a map displaying women ratio by seat. I used similar technique as before to generate the map relying on Geo-JSON data it is worth however mentioning that all National Assembly elections for each seat were group together and averaged out to generate a mean average display of each province which allows some outlier data to stand out and less variation among certain points. This was done by grouping together the constituency name across the multiple elections and taking the average followed by using the most recent geographical data of that seat.

The last few segments of research was focused on the male and female outliers which was determined to be so if its z-score was greater than 2 standard deviations away from mean. Any data-point two standard deviations below mean where there was an abnormally large amount of men voting was classified as a male outlier where 2 standard deviations above classified a female outlier.

These points could then similar to previous map be plotted using a scatter plot. This form of plot was chosen as it allowed data to not be covered up and visible to reader. We also identified the correlations using the matrix strategy from before between gender outliers and various available features.

III. RESULTS

A. Party Proportions

Starting off in the analysis of Pakistani politics is the following question. Who are the major Pakistani Parties?

National Assembly Seat Distribution by Political Party in Pakistan

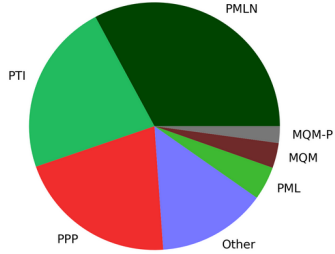


Figure 1. Seat Distribution by Party

Going off figure 1 we can clearly see that Pakistan is multi-party system. 3 Major parties the PML-N, PTI, and PPP make up the largest major parties. However, Pakistan also contains a considerable number of Independents and small minor parties lumped into the other category.

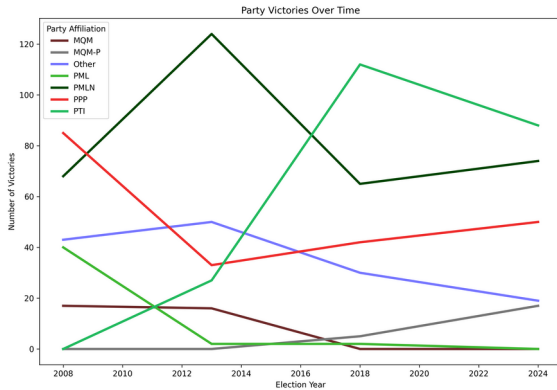


Figure 2. Number of Seats Won Over Time

Using figure 2 we can continue these results and see the trajectories and changes of these major political parties the most noticeable change is the rapid and recent rise of the PTI from a non-existent party to the largest political party over the course of the last 3 elections. Alongside this we also seem a decline in the number of seats won the PML-N particularly in the 2018 where they dropped from a strong first to a weaker second. Thirdly, in regards to the PPP we see an initial loss in seats in 2013 however a slow and but steady increase over the following 3 elections. Lastly, the most significant decline has been in the other section of parties whom have consistently

lost seats across the last 3 elections and seem to be on a downwards trend.

B. Geographical Dominance of Parties

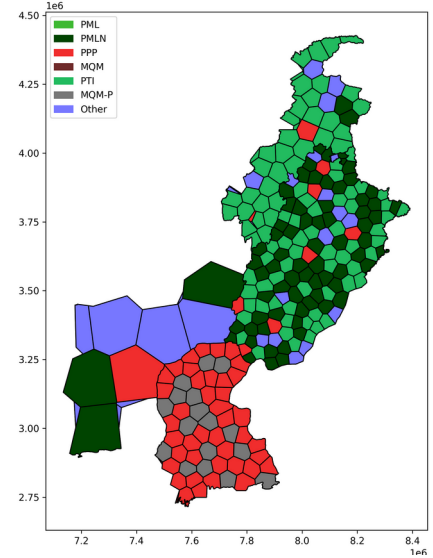


Figure 3. Geographical Location of Seats

Going into some of the regional areas of specific part given by Figure 3 we can see most noticeably is a stark divide between seats controlled by PPP and the other major parties where PPP dominance is deeply concentrated within the southern regions of Pakistan. We also see a lighter divide between PTI and PML-N in the northern regions of Pakistan. Generally, PTI controls most seats in the north western region while PML-N controls the North eastern region.

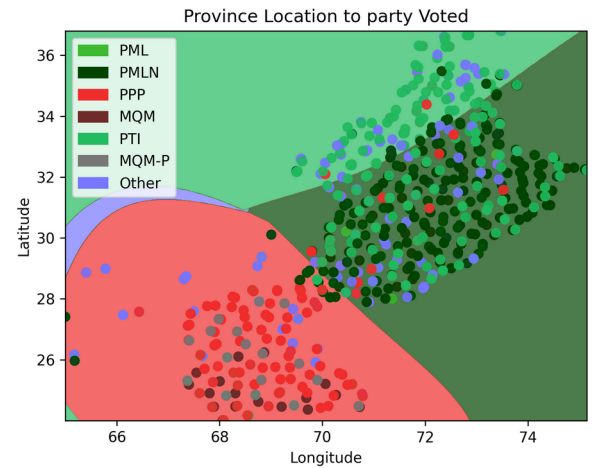


Figure 4. Geographical Scatter-plot

Figure 4 further confirms many of the ideas as the SVC model too picks up those same regional dominance's as described. On top of this we can confirm that there is a degree

of legitimacy among this as accuracy of the model is at 51.6% a roughly 19% increase from the dummy classifier's accuracy of 32.8%. On top of this the model shows the potential for a fourth unnoticed region of small parties which are popular among the south western border of Pakistan. However, this is more nonconclusive as the regression model was unable to pick up any points in this region.

C. Women Ratio Distribution Curve

Using the complete dataset we have have that the mean ratio of a seat whos voters are women is 44.2% while the median value is 44.8%. Due to the similarity between these two values it is reasonable to assume that that his a fairly non-skewed dataset.

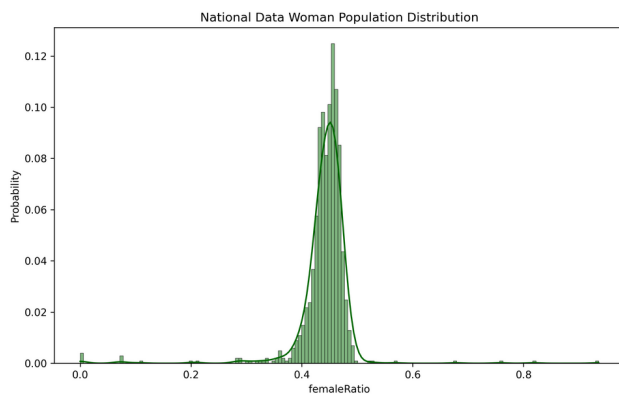


Figure 5. Female Voting Ratio Distribution Curve

Viewing the curve given in figure 5 it continues to suggest that the data is in-fact not skewed and that it also may in fact be a normal distribution

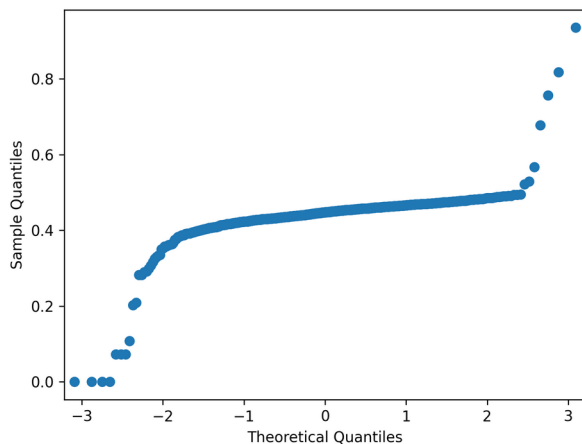


Figure 6. Female Voting Ratio Distribution QQ-Plot

We can see however from figure 6. an alternative conclusion that the dataset is in fact not normal due to the non-linearity of the plot seemingly caused by a set of outlying

points. This hypothesis is further shown when using the Shapiro-Wilkes hypothesis test on the dataset we get a p-value of 0.00 which indicates a strong rejection of the null hypothesis of the dataset being normal and instead conclude that this is in-fact not normal dataset.

D. Women Proportion by Party

Following the goal of answering the previous given questions it is logical to analyze the relationship between women and the party they voted for. Although this data is not explicitly given we have instead chosen to see relation between high ratio of women voters and the party elected.

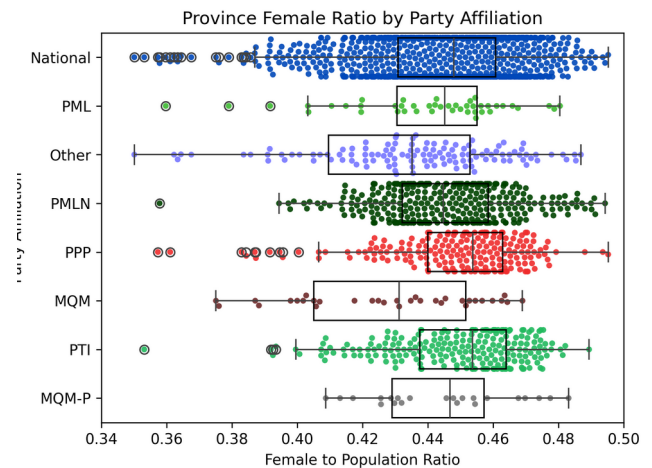


Figure 7. Close-up Female Ratio by Party

Analyzing the graph in figure 7 we know the following information. It appears that the PTI and PPP are the only two major parties who contain a higher than national average percentage of women in the constituencies they won. We can also see large variations in certain parties such as Other, MQM, and PPP as they stretch greater distances. Using hypothesis testing we were able to conclude that with a alpha value of .05 the following hypothesis were accepted. Firstly, constituencies with higher than average women populations are more likely to vote for PPP and PTI. Secondly, MQM, Small Minor parties are less likely to win seats with higher women ratio and Lastly, PML-N, PML, and MQM-P null hypothesis of lower than national average couldn't be rejected. The p-values of these parties were as followed .0008, .0009, .0006, $2 * 10^{-7}$, .69, .15, .64

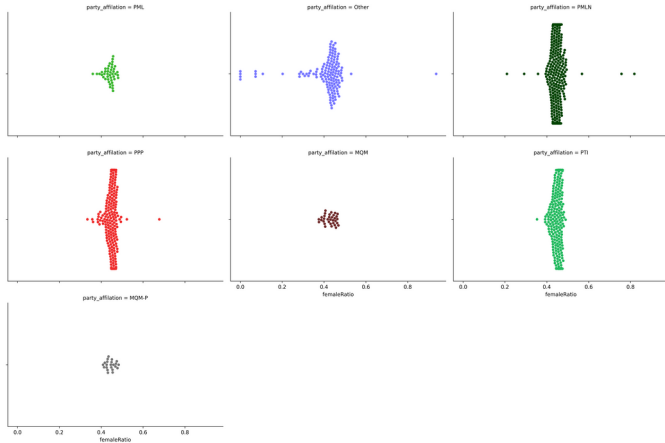


Figure 8. Full-view women ratio by party

It is worth taking notice however that figure was misleading in showing how close the points were to each other and in fact as seen in figure 6. Pakistan contains many outlying data points. These data points seemed to be strongest in the Other section with a couple in PML-N and PPP.

E. Gender Correlations

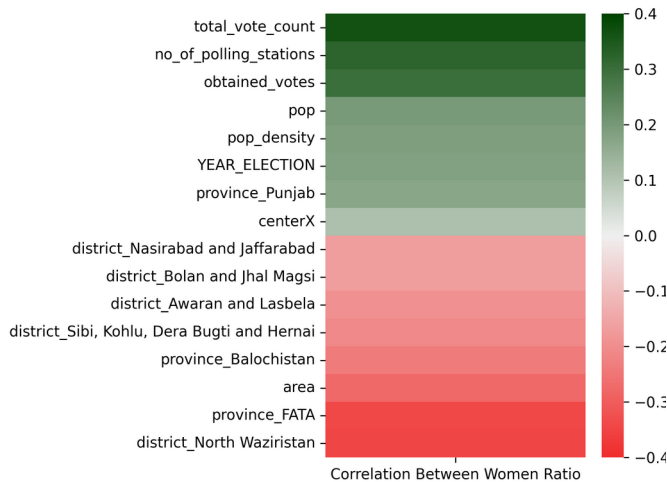


Figure 9. Women Voting Ratio Correlation Vector

Going off of figure 9 we can see that the ratio of women to men to vote is highest in areas of high population and various statistics that relate to number of people such as total number of votes, population density, and obtained votes. We can also see that the ratio if women to men has increased over the past 4 elections. However, on the other side we see that there is negative correlations between area and specific regions of Pakistan. Most notably is the province of FATA or Federally Administered Tribal Areas one of the special regions in Pakistan.

Using the other features in the dataset I also calculated the percentage of voters who are women which is one of the most

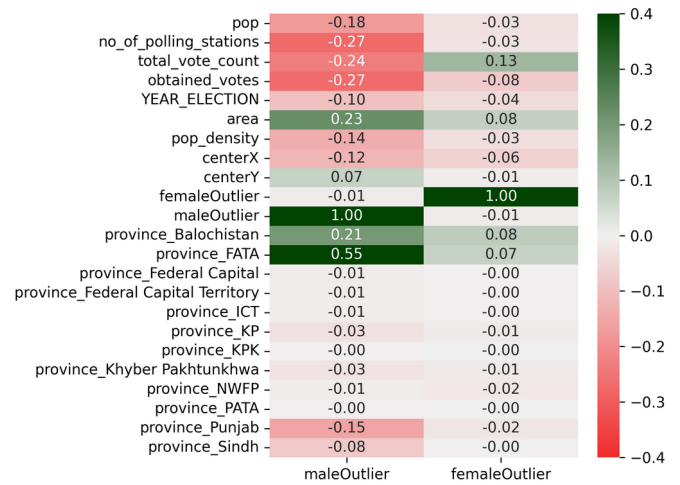


Figure 10. Gender Outlier Correlation Matrix

Analyzing the male outliers in figure 10 continue to see a trend that is inverted from the one seen figure 9. This is a logical continuation of the data and likely reaffirms that the correlations being made have some legitimacy. Interestingly data in the female outlier sections however have significantly less correlation with the exception of total vote count we also lines up from results in figure 9.

F. Outlier Locations

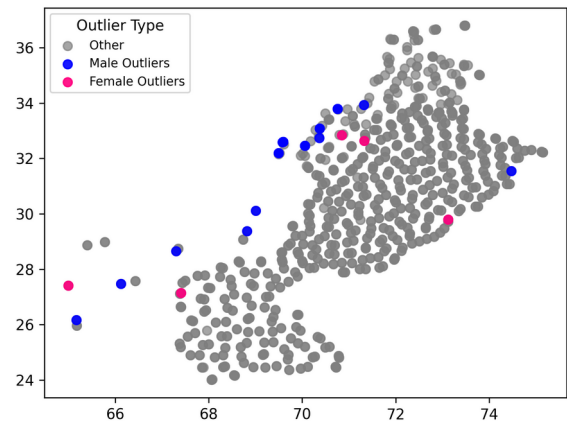


Figure 11. Outlier Location Scatterplot

Looking at figure 11 we can see a trend similar to the one given by the regression in figure 4 where there exists a an outside rim of unusual results. In this case we can see that male outliers have the tendency existing in the western edges of Pakistan while most female outliers are deeper into Pakistan's interior region.

central components to research. The Z-score of that ratio and a true or false boolean category that holds whether to consider

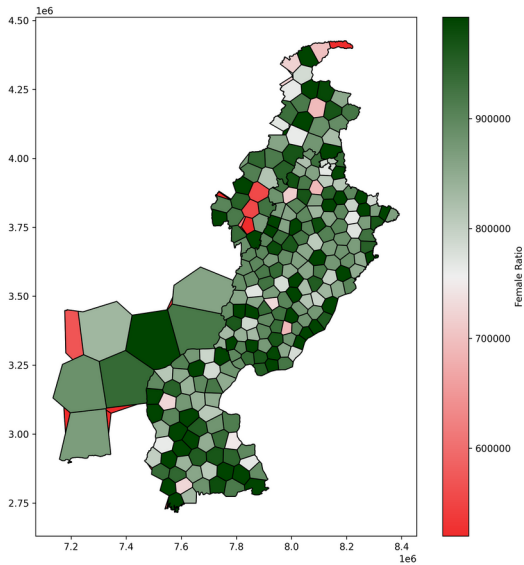


Figure 12. Average Female Ratio by Constituency map

With figure 12, we can get a more generalized view of the results of figure 11, where we do in fact once again see that the male outlying constituencies do seem to all exist and that many of the female constituencies once again as shown in figure 11 exist more eastwards.

IV. CONCLUSION

To answer the first question Who are the major political parties in Pakistan? We can work off figure 1, which indicates that PML-N, PPP, and PTI are by far the largest political parties in Pakistan however there still lies a great number of seats in the other smaller parties along with independents that make up the rest of the Pakistani electorate. It also becomes obvious from this result that Pakistan is a multi-party system.

Secondly, How have these party dichotomies changed over time? This question can be simply answered using figure 2 which gives that we have seen a quick and rapid growth in the popularity in PTI within recent rising from a party of near non existence to the plurality of the National Assembly. We can also see the influence of the PML-N particularly in recent elections beginning to dwindle along with the increase in seats by PTI. Next, although a immediate decline initially we have seen that across the following 3 elections an slower increase and stabilization of number of seats being held by the PPP. Lastly, we can see that the number of seats held by other smaller parties and independents has been on a rapid decline in these recent elections.

Next, the third question asked is what factors are correlated with a high ratio of voters being women? This question is best answered through figure 9 & figure 10 both of these figures suggest that there is a strong positive correlation between various population related statistics such as

population, population density, number of votes, obtained votes, and number of polling stations all of which are built on the number of people who exist within a constituency. We also have seen positive correlation between women voting ratio and elections that have occurred more recently meaning the mean ratio of women voters has been on incline and assuming this correlation holds would mean that the mean ratio of women in a constituency should increase over time indicating a increasingly more relevant women voting base. Lastly, we can see off these figures that low women ratios seem to have the highest concentration with specific districts and provinces most notably the province of FATA being the only province shown with a strong negative correlation.

Lastly, for the final question: How might the women voters be influencing Pakistani party dichotomies? This question is by far the most complex of the 3 however this research can hint at some possible ways that women might be influencing. Firstly going off of our previous results we can see that number of women voters has been on the incline over time. On top of this parties such as PPP and PTI over the last few elections have both seen an increase in number of seats and based on results in hypothesis testing and results from figure 7 we can also see that these two parties are also strongly correlated with high women constituencies. On top of this Independents and small parties who tend to be popular in very low female ratio provinces have seen significant decline across recent elections.

If one were to continue this research the area in most need of continuation would be in regards to women's influence on party dichotomies. Factors that could be influencing results and not tested in this research would be attempts to try and to show that the affect that women have on winning parties isolated from the various correlations such as high population and location as these indicators could be giving misleading information in regards to influence of women. More research can also be put into a deeper historical analysis to see if women votes have been increasing over a longer period of time and if so if results similar to the ones provided here where parties that dominate in high woman ratio constituencies tend to see increases in party performance in these elections.

REFERENCES

- [1] Pakistan - The World Factbook. (n.d.). Wwww.cia.gov. <https://www.cia.gov/the-world-factbook/countries/pakistan/#government>
- [2] GAHATRAJ, D. (n.d.). SUBJECT: POLITICAL SCIENCE VI COURSE: BA LLB SEMESTER V (NON-CBCS). https://www.iilsindia.com/study-material/743837_1610305745.pdf
- [3] Siddiqui, N., Mufti, M., & Sahar Shafqat. (2020). Introduction: Pakistan's political parties in an era of transition. In *Pakistan's Political Parties: Surviving between Dictatorship and Democracy* (pp. 1-20). Georgetown University Press. <http://www.jstor.org/stable/j.ctv10crrkk.6>
- [4] Raza, M. (2024). *Pakistan Elections Data 2024* [Data set]. Kaggle. Retrieved from <https://www.kaggle.com/datasets/mesumraza/pakistan-elections-data-2024/data>
- [5] OpenAI. (2024). *ChatGPT* (November 20 version) [Large language model]. Retrieved from <https://chat.openai.com/>

