Political Polarization and Marriage

Final Project Report

Department of Computer Science, Stony Brook University

Abstract – Aim to analyze how there has been an increase in political polarization in the United States and how it affects the cohabiting of married couple. By extensive data analysis it should be possible to identify cohabiting couples and track them over each electoral snapshot to see whether they are still cohabiting. We are looking at the possibility of whether the couples have ceased cohabiting due to political polarization.

1. INTRODUCTION

America is a divided society. The most significant fault line in the second decade of the 21st Century is not race, religion or economic status, but political party affiliation [1]. A decade ago, the public was less ideologically consistent than it is today.

In 2004, only about one-in-ten Americans were uniformly liberal or conservative across most values. Today, the share who are ideologically consistent has doubled: 21% express either consistently liberal or conservative opinions across a range of issues – the size and scope of government, the environment, foreign policy and many others [2].

Effects of this antipathy among people with dissimilar affiliations, can be seen in various social settings. One such social setting discussed in one of the papers is about thanksgiving dinner. They show that Thanksgiving dinners attended by residents from opposing-party precincts were 30 to 50 minutes shorter than same-party dinners. Moreover, the effect is tripled in the year 2016 when the attending people belonged to media markets with heavy political advertising. Moving the analysis to a more private setting, we are analyzing the effects of diverse political beliefs between couples.

Similar analysis titled "How Many Republicans Marry Democrats" showed few interesting results. Firstly, 30 percent of married couples have dissimilar political affiliation. A third of those are Democrats married to Republicans. The others are partisans married to independents. Also, in the Democratic-Republican pairs the male partner being republican is twice as many as female partner being republican. Second, 55 percent of married couples are Democratic-only or Republican-only. Third, there is a much higher rate of mixed-partisan couples among younger pairs than older pairs [3].

Similarly, another paper suggested how political polarization has impacted the family ties of the people. It works on the same situation and that is the Thanksgiving Dinner. The data was got from anonymized smart-phones and precinct-level. Few methods were used to predict the duration of the Thanksgiving dinner of couple with different party affiliations Our scope of study here was to find the effect on marriage and how does it provide us evidence that they are not staying together(For example – Different address of the people who were once together). The results show that the dinner duration reduced by 30-50 min on average. Also, the Republicans were more sensitive and reduced their dinner duration by more minutes than their Democratic counterparts.[4]

Another paper also sights interesting insight on the dating life of the people. Although this paper is more about how people date the guy/girl based on the political preference, we get a sense as to how Political Polarization plays a significant role when it comes to couples staying in the United States.[5]

All these papers give us an idea as to how Political Polarization is one of the factors which affect the cohabitating of married couple. Soon with the help of the Florida County data we will try to make some analysis and quantify the effects of Political Polarization.

2. DATASET

We have generated a dataset using the data from Florida Voter Registration System

The Florida Voter Registration gives us the data over its 67 counties over the years (2012 to 2019). It helps us to find out many things about the people living in each county. These are our features which we use to find out a trend for the people staying in Florida.

The data which we are working on is got from Florida registered voter data. It contains data extracted from the Florida Voter Registration System. It includes information on voters who are officially registered.

So, the fields which we are using based on the Voter Registration Data is:

- **Voter ID** – A 9-digit ID which is unique to all the residents of Florida irrespective of the county in which they stay in.
- **County Code** – As there are 67 counties in Florida State, each county is represented by a unique 3 letter County Code. (Example – POL for POLK county)
- **Names** – First Name, Last Name, Middle Name, and even Suffix data is present.

- **Residence Address and Mailing Address**- Full Address of the residents.
- **Gender** and **Race** of each resident.
- **Birthdate**
- **Party Affiliation** of each resident.
- **Sex** of each resident.
- **Age –** This is found out based on the Birthdate and the current year for each resident.
- **Voter Status** – This tells us whether the resident is eligible to vote.

2.1 DATA PRE-PROCESSING

The information of the people who have voted is available to the public except those cases where an exemption has been applied. The dataset includes features like First Name, Middle Name, Last Name, Mailing and Residence Address, Gender, Race Party Affiliation, Voter Status, Birth Date and so on.

There are many datasets corresponding to different county and different years. So, we need to check for the trend in the vote, mailing address and the party affiliation of opposite gender. This will give us an idea about the polarization when it comes to marriage and political party. So, we have now found out the Voter ID of the spouse who has the same Resident Address and the age difference between them is less than or equal to 10 years. This data is spread over the years, that is from 2012 to 2019. This gives us the trend of Party Affiliation of each user.

We also have the voting history from 67 county supervisors of elections. The data is captured at a point in time and may not always correspond to information in a county's voter registration file because the voter may have moved to another county after voting.

As mentioned above, we plan to merge the data belonging to the same county over different years and the voting history which we get every year. This data is got from the Registered Florida User Data.

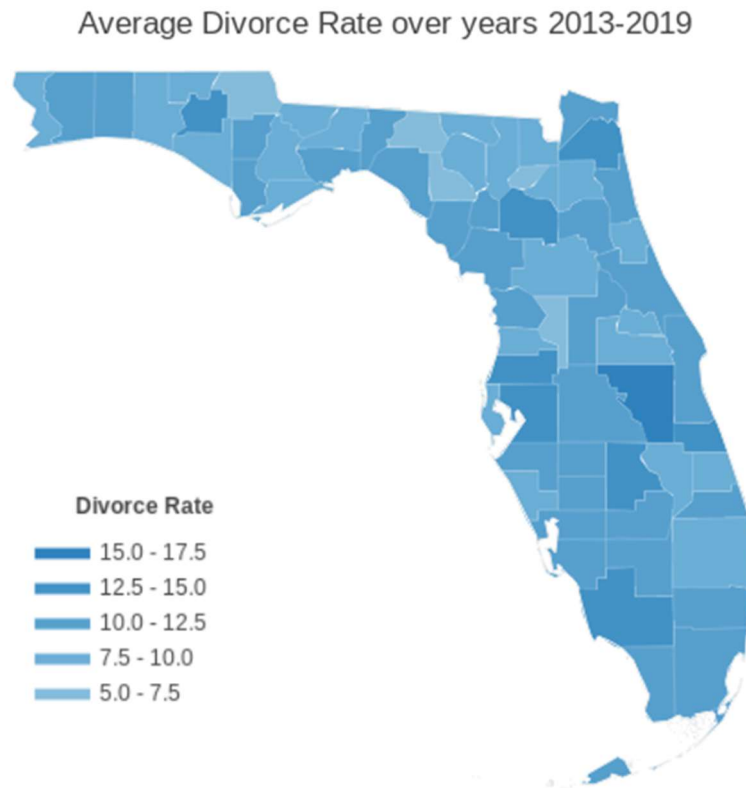Average Divorce Rate over years 2013-2019

Fig 1: Map that shows the Average Divorce Rate over years 2013-2019

After merging the datasets, we then check for the trend whether 2 people of opposite gender get married based on the mailing address change or there is a change in the Last Name. After finding the couple who are married, we check for the political affiliation for each of them. If there is a trend in the votes of the political affiliation has changed post marriage or is it still the same.

3. METHODS

Approach 1: *Distributed Computing and Merged Analysis Over years*

1) This was one of our first approaches wherein all the datasets belonging to various counties and different counties were individually cleaned.

2)Then data of all counties was merged for a particular year and then data over the years were merged before computation.

Drawback:

As the dataset was large, we realized this was not a feasible approach and merging all dataset was very inefficient way of analysis.

Approach 2: Distributed Computing and Merged Analysis Over Counties

1)Here the data was cleaned and massaged as to obtain a dataset from which meaningful data can be extracted easily.

2) The was so arranged that data of a county had all the information needed for polarization analysis over the years. Basically, the data over

the years and quarter year data over the years 2012 to 2019 was merged get a master dataset for each county.

3) The necessary computation was performed on the above merged chunks to get a meaningful compressed data.

4) Multiple Counties, each containing the data over years for the county were merged to observe the effect of the party affiliation on politics

Drawback:

Even though the data analyzed provided meaningful insight of the trend of divorces, party affiliation over years, it was hard to continue with this approach due to its inefficient way of analyzing data, resulting in generation of data as huge as 2GB when 6 counties were merged. Even the time consumed for analysis made it difficult to move forward with the approach.

Approach 3: Distributed Computing and Distributed Analysis Over Counties

After a thorough analysis of data, we realized that the analysis could be performed by grouping the data and analyzing them separately and finally merge the results.

1) We observed that when we consider one county over the years, we get some meaningful results for each year and then this can help us find the total number of couples who got married or divorced or left the county in a particular year.

2) The above result was got by performing analysis over the data merged over the years for each county.

3) This made it quite easy for us to classify each couple who is married or divorced along with their Party Affiliation.

4) This drastically reduced the size of the dataset (maximum size of 500Mb) under consideration.

5) The dataset under consideration remained almost the same in compared to approach 2 but a huge difference was seen in the ease of computation when this data was analyzed as it is without merging.

6) So we were able to obtain all the necessary results for each county and cumulative of these results was obtained to arrive at the expected analysis.

3.1 STEP BY STEP APPROACH

1. The Distributed Computing and Distributed Analysis Over Counties had more benefits with the ease of arriving at the result with almost the same result as the Distributed Computing and Merged Analysis Over Counties approach so we continued with it.
2. We first considered county by county data across Florida. Each county had the information on voters over the years (2012 to 2019).
3. When we consider one county, we look for 3 aspects which we feel are quite important in order to find out more about the Political Polarization.
   i. We check if there are any people who got married in that particular year. We got to know if they got married in a particular year say "X" if they have the same Residence Address in year "X" and their Residential address differed in the previous year "X-1". Also, we considered the age difference a filtering

parameter when we arrived at the result.

ii.   We also need to check if there are any cases of divorce in that particular year. We get this information if 2 Voter-Id which had the same Residence Address are now different. As in addresses were same in "X-1" and became different in "X"

iii. We considered the people who have left the county in a particular year as the Voting Information from that particular year is Null in the county where they previously lived.

4. These features accompanied with the verified assumptions formed the basis the results obtained.

5. 5. This data is collected over the span of years for each county to find the divorce trends, party affiliation trends, marriage trends in counties

6. We found out the count, weighted average and percentage metrics for each county which formed the results of divorce, marriage, percent divorced, party affiliation change.

7. The individual results were merged by using summing the counts and performing a weighted percentage of all the counties giving us the overall trend in Florida.

Example of one such analysis by cumulative result was the divorce trend. We found the

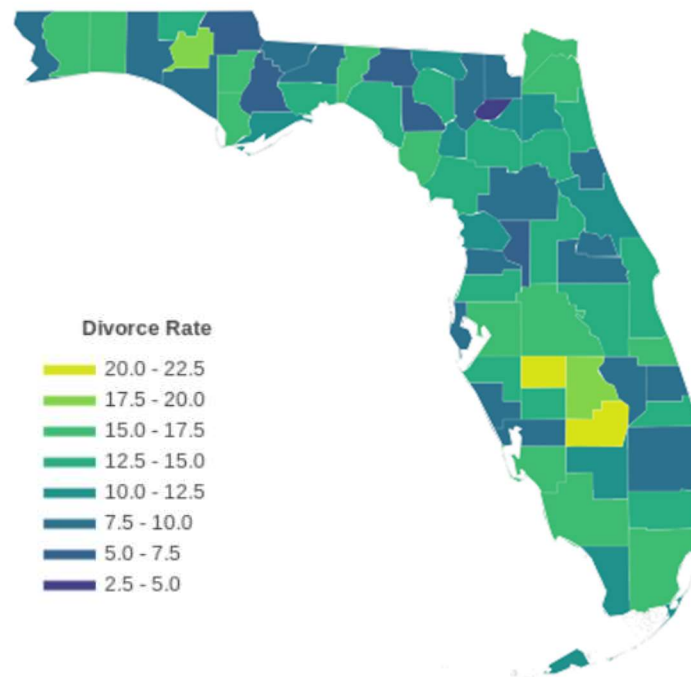Fig 2: Map that shows the Divorce Rate of couples with different political affiliations before 2016

Fig 3: Map that shows the Divorce Rate of couples with different political affiliations after 2016 elections

number of people getting divorced having same and different party affiliation in each county in every year. If "Y1" is the county we found out the number of people having the same party affiliation getting divorced in "Y1" in years 2012 through 2019 say "S1[2012...2019]" and similar count of people having different party affiliation and getting divorced say "D1[2012...2019]". Similarly, we found the counts for all the 60 odd counties i.e., S1...Sn [2012...2019] and D1...Dn [2012...2019]. Summation of these values and ratio over the summation provided the influence of party affiliation on divorces.

RESULTS:

Here we have considered the divorce rate before and after 2016 elections. The first figure gives us an idea about the average divorce rate over the year 2013- 2016 for different counties. We iterated over the counties to get the number of divorces where in the party affiliation amongst

the couple was different for each year that is from 2013 to 2019. So, once we get the count of divorced couples over all the counties and each year, we segregate the result into 2 parts i.e before 2016 elections and after 2016 elections. The first graph depicts an average of the divorced rate of all the years before 2016 elections in various counties of Florida. Similarly, the second graph represents the average divorced rate over different counties after the 2016 elections. We observe that there has been a slight increase in the Divorced rate after 2016 election. This was the when Donald Trump was elected the President of the United States. As each of the scales in the graphs are same, the Florida map gives a better representation of the increase in the divorce rate. There has been an increase of **26%** when we consider the average divorce rate of Florida State before and after 2016 elections.

Percentages	County Codes
-50 to -40	LAF
-30 to -20	BRA, CHA, GAD
-20 to -10	BAK, HOL
-10 to 0	CIL, DIX, HER, LEV, VOL
0 to 10	ALA, DES, ESC, FRA, MON, MRM, OKA, OKE, OSC, PAL, PAS, PUT, SAR, SUM, WAL
10 to 20	BAY, CAL, GUL, IND, LEO, ORA, PIN, STL, UNI, WAS
20 to 30	CLA, FLA, GLA, HIL, LAK, MAN, MRT, STJ
3o to 40	CLM, GIL, HEN, SEM, SUW
40 to 50	DAD, DUV, JSC, NAS, SAN, TAY

Table 1: Table of Percentage change of divorce rate before and after 2016 elections for different counties

This Table gives us a clear picture as to how the there has been an increase in the divorce rate percentage over different counties. The minus percentages indicate that there are few counties where the divorce rate was less compared to the previous year i.e. 2016. There are quite a few counties who have such decreasing divorce rate.

However, when we see the positive percentage change i.e. from 0-10,20-30 and so on we can see that there has been an increase in percentage of divorce rate and this table above represents the percentage as to how much has changes post 2016.

3.2 ALGORITHM

Algorithms 1 and 2 were basically used for cleaning and massaging the datasets individually so that meaningful data can be extracted.

Algorithm 1 runs over all the counties and all the years cleaning and merging the counties data by years. So, all the necessary columns are so merged as to obtain columns of these over years. For example, party affiliation is one of the

Algorithm 1: Algorithm

```
for i in counties: do
    p = initialise i^th county's dataframe;
    for j in years: do
        p.clean() and p.merge( i^th county's j^th year );
    end
    write_to_csv(p);
end
```

Algorithm 2: Algorithm

```
for i in counties: do
    county_data = read the merged county datafile;
    for j in years: do
        county_data = append is_divorced, is_married, is_couple columns of j^th year to data ;
        county_data = append couplePartyAffiliationCombination column;
    end
    write_to_csv(county_data);
end
```

Algorithm 3: Algorithm

```
metric[1] = {'1^st year':0, ....,'n^th year':0};
.
.
metric[N] = {'1^st':0, ....,'n^th year':0} ;
for i in counties: do
    county_data = read(i^th county's data);
    for j in years: do
        metric[1][j] = {'1^st year':0, ....,'n^th year':j};
        .
        .
        metric[N][j] = {'1^st':0, ....,'n^th year':j} ;
    end
    for i in years: do
        obtain results from metric[i];
    end
end
```

columns of our interest. The merged data frame for county say 'A' has columns 2012 party affiliation, 2013 party affiliation ad so on till 2019 when it is passed through algorithm 1. Similarly, all the county is passed through algorithm 1 and the merged master data is written back to drive so that computation can be performed in the space and time constraint environment.

So, the columns obtained till now were a merge over the years. So, extract information from these columns algorithm 2 is used. Here we create new columns which can be used to classify people. Few examples, of it is to map the couple together based on residential address and age, extract age based on current year and birth date, checking if the people are divorced

based on the residential address in consecutive years. So, a few such data columns were added by passing them through the algorithm 2.

The algorithm 3 was used to get a cumulative analysis over the counties. So, the Algorithm 3 generates count, averages on top of the columns generated in algorithm 2. A cumulation of these counts and averages over the counties provided the necessary results. An example of one such computation was the divorce rate. The count of the column 'is divorced' which was generated by algorithm 2 provided with the count of people getting divorced in each year in every county. And the cumulation over counties provided with the number of people getting divorced in each year over entire state. Many such results were obtained by passing the results of algorithm 2 through algorithm 3.
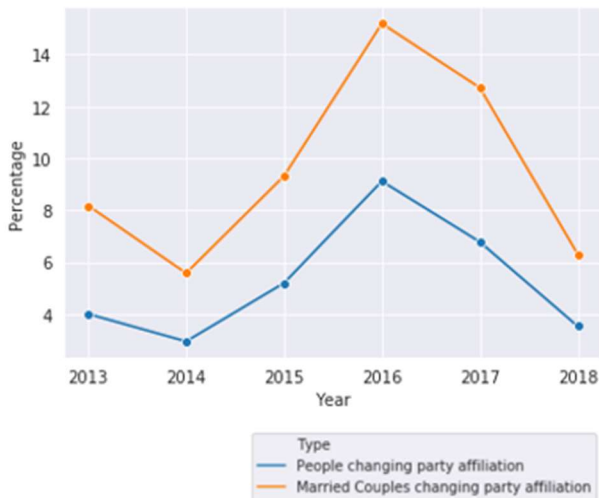

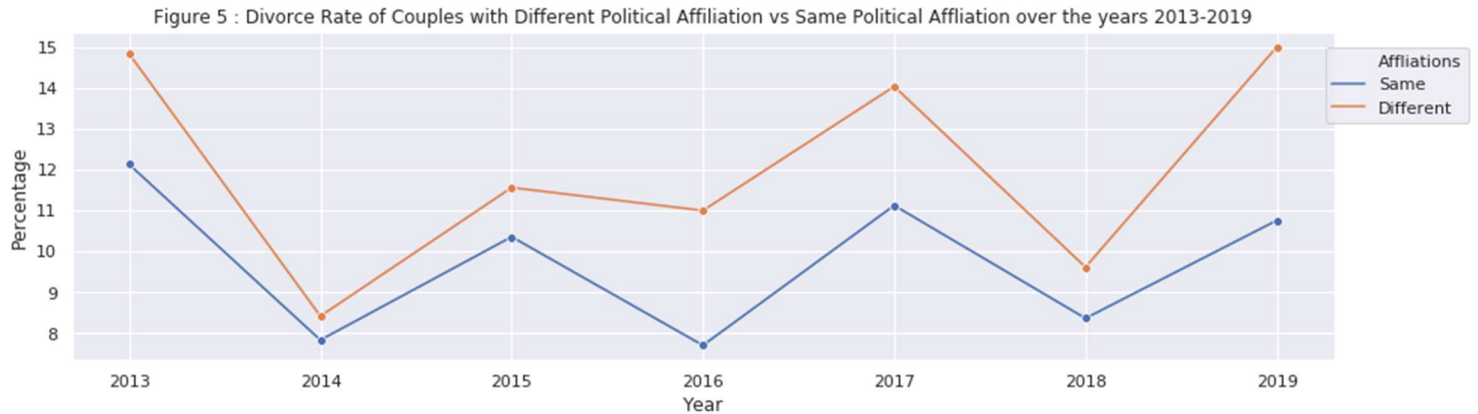Fig:4 Convergence of Political Affiliations

4. RESULTS

- With the help of **Fig 4** we can infer something about the effect of marriage on party affiliation. This is our area of interest as it gives a clear evidence as to

Political Polarization has some effect on the couples. So, in order to get this metric, we find out the count of people who change their party affiliation through the years 2012-2019 in every county. This process is then repeated to find out the number of party affiliation changes in couple alone. In order to find out if who the couples are, we use their respective Age and the residential address as discussed in the approach.

- Once we get the total number of couples and individuals changing their party affiliation, we aggregate these counts over all the counties in Florida state.
- So as seen in the graph, each year we take the percentage of people and couples changing their party affiliation and plot it over different years.
- It is evident that there has been a significant increase in the percentage of couples as well as individuals changing their party affiliation in the year 2016. AS discussed in Figure 1 and 2 above, the results of election 2016 have a strong effect on the people.
- The main intention as to why we plotted this graph is to prove how Political Polarizations plays an important role in the lives of couple. From the graph above, party affiliation changes are quite frequent when it comes to married people rather than individuals.

Fig 5 gives us an idea about the Divorce Rate of the couples who have different affiliations amongst each other and the couples with same party affiliations.

As we find out the married couples, we find out the total number of divorces happening every year from 2013 to 2019 over different counties. We keep a track of both the count over the period of years.

Figure 5 : Divorce Rate of Couples with Different Political Affiliation vs Same Political Affliation over the years 2013-2019

Once we find out the count of divorces of couples with same and different party affiliation, we find out the divorce rates for both the entities depending on the total count of couples.

This statistic is then taken for every year and plotted to get a trend. This trend tells us that the divorce rate is quite high when the couples have different party affiliations when compared to the couples who support the same party. This is quite a concrete evidence to believe that political polarization does play an important role when it comes to couple who have ceased cohabitating with each other.

represents Republic Party. So, by DEM-DEM we mean in a couple the man supports Democratic party and the women supports Democratic Party too.

Similarly, when we consider DEM-REP we mean the man supports Democratic Party and women is more inclined towards Republican Party. Like these 2 above mentioned combinations we have REP-REP and REP-DEM too.

Figure 6. : Divorce Rate of all party affiliation permutations
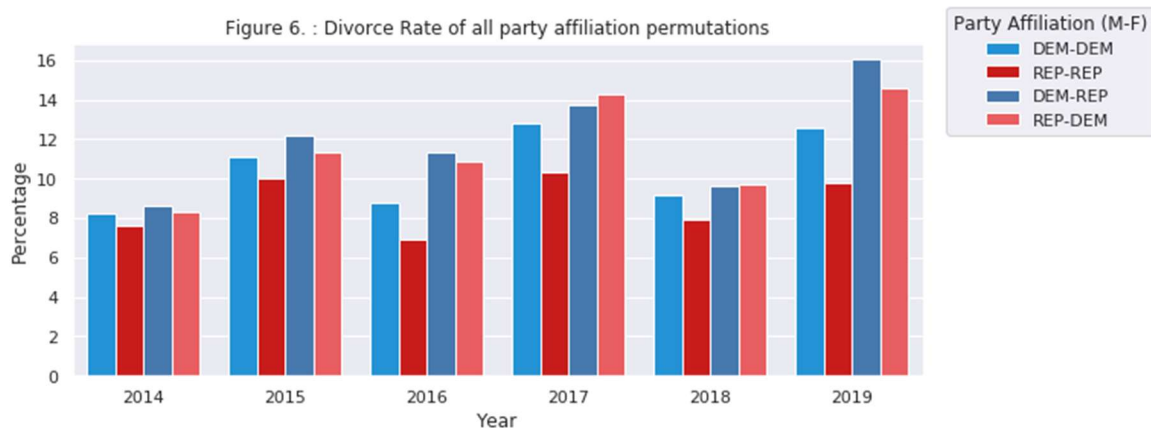
The **Fig 6** gives us the total count of couples who are divorced along with their current political party affiliations for that particular year. Here DEM represents Democratic Party and REP

5. CHALLENGES FACED
- One of the challenges we faced was large dataset which was solved using the Distributed Computing and Distributed

Analysis Over Counties Approach as discussed above.

- Since the computation and analysis were performed for each county and this resulted in performing same steps repeatedly. Batch operations performing the steps iterating over counties, working on single county at any point of time eased out the issue.
- With the improved approach the data size was in good shape, but the computation still took few hours and often a recompilation was needed in case of mistake. So faster techniques in all places possible. (Like Pandas merge, NumPy vector fared better over lambdas.)
- The major concern was regarding the data of the people who were migrating to different counties and their influence of the overall outcome. But after a thorough analysis we found that our result was tolerant to migration as it

So, if people migrated from county "A" to county "B" and "C" in year "X". Then in year "X+1", the data of migrated people was considered in county "B" and "C" instead of "A". This way the data was people was a part of at least one of the counties in every year. The data of our interest was (A U B U C) which covered the migration concern.

6. CONCLUSION

1. We observed that the rate of divorces is slightly higher among the couples with different party affiliate (Example husband DEM supporter and wife REP supporter)

2. One other observation was that few people change the party affiliation on marriage which is presumably the influence of the spouse's party affiliation.

3. The divorces among both category of couples- couples with same party affiliation and couples with different party affiliation increased after 2016 election compared to before election.

4. The overall divorce rate varied from as low as 8% in 2014 to 15% in 2019 with an overall of 12.38% which in comparison with the observed 13.0% from USCToday[9].

7. REFRENCES

[1] https://pcl.stanford.edu/research/2017/iyengar-home-political-fortress.pdf
[2] https://www.peoplepress.org/2014/06/2/section-1-growing-ideological-consistency/
[3] https://fivethirtyeight.com/features/how-many-republicans-marry-democrats/
[4] Mixed partisan households and electoral participation in the United States by Eitan HershID, Yair Ghitza
[5] New Wakefield Research Study: The Trump Effect on American Relationships.
[6] Dataset for map of Florida- https://catalog.data.gov/dataset/tiger-line-shapefile-2016-state-florida-current-county-subdivision-state-based
[7] For Map Plots - https://plot.ly/python/county-choropleth/
[8] For Numpy Vector - https://towardsdatascience.com/how-to-make-your-pandas-loop-71-803-times-faster-805030df4f06
[9] https://www.usatoday.com/story/money/economy/2018/02/02/broken-hearts-rundown-divorce-capital-every-state/1078283001/