Capstone Project - Milestone Report

Lily McElwee

**INTRODUCTION**

My project studies shifting political sentiment in Silicon Valley using data on individual contributions to presidential election campaigns in the 2008, 2012, and 2016 cycles. Historically, California is a "blue"" state; in the 2015-16 cycle to date, the Democratic party and candidates have raked in 58.4% of total contributions while the Republicans have received just 37.7%. But while Silicon Valley has leaned left in every presidential election going since 1984, data through the end of 2015 shows that contributions to Republicans from employees of several major tech firms (Cisco, Oracle, Yahoo, Intel) has so far outstripped money flowing to Democrats in the 2016 cycle. Indeed, while just 52 workers from the tech industry have contributed to Trump’s campaign specifically, analysis of past election cycles indicates Valley support for Republicans overall is not unsubstantial (<http://www.bloomberg.com/politics/articles/2016-02-09/bush-rubio-lure-tech-geek-giving-googler-dollars-back-clinton>) and may be growing.

For the purposes of framing the research, my project takes as its client the California Republican Party. Faced with an uphill battle, leaders from the party focused on Silicon Valley want to improve fundraising efforts by directing marketing expenditures to those areas in which they have, on average over the past three presidential election cycles, sourced the lowest amounts of funding and for which funding declined in 2016 (to date) versus 2008.

Articles on the volume (ibtimes.com) and breakdown of individual contributions abound in the popular press, but most examine presidential campaign funding in light of data from a single quarter or cycle. Silicon Valley is demographically dynamic and famously unideological, and fully understanding the opportunities for a specific political party means examining sentiment in recent historical context. In this project, I look at data over the past three election cycles to extract overall and zipcode-specific trends, and identify those areas in which funding to Republicans has a) declined (perhaps indicating an area in which increased marketing focus is needed) and b) risen as a share of total funding (perhaps indicating emerging Republican strongholds worth continued marketing efforts).

**DATA**

Background

I use the FEC’s official data on presidential campaign contributions from a list of Silicon Valley zip codes for 2008, 2012, and 2016 election cycles. The data can be found by entering specific zip codes here: [http://www.fec.gov/disclosurep/pnational.do#](http://www.fec.gov/disclosurep/pnational.do). I have identified 63 zip codes comprising the “Silicon Valley” area, summarized in the table below.

## County Zipcode  
## 1 Campbell  95008   
## 2 Cupertino 95014  
## 3 Gilroy 95020  
## 4 Los Altos 94022  
## 5 Los Altos Hills 94024  
## 6 Los Gatos 95030  
## 7 Los Gatos 95032  
## 8 Los Gatos Mtns 95033  
## 9 Milpitas 95035  
## 10 Monte Sereno 95030  
## 11 Morgan Hill 95037  
## 12 Mountain View 94040  
## 13 Mountain View 94041  
## 14 Mountain View 94043  
## 15 Palo Alto 94301  
## 16 Palo Alto 94303  
## 17 Palo Alto 94304  
## 18 Palo Alto 94306  
## 19 Palo Alto 94309  
## 20 Palo Alto 94302  
## 21 Palo Alto 94305  
## 22 San Jose 95110  
## 23 San Jose 95111  
## 24 San Jose 95112  
## 25 San Jose 95113  
## 26 San Jose 95114  
## 27 San Jose 95115  
## 28 San Jose 95116  
## 29 San Jose 95117  
## 30 San Jose 95118  
## 31 San Jose 95119  
## 32 San Jose 95120  
## 33 San Jose 95121  
## 34 San Jose 95122  
## 35 San Jose 95123  
## 36 San Jose 95124  
## 37 San Jose 95125  
## 38 San Jose 95126  
## 39 San Jose 95127  
## 40 San Jose 95128  
## 41 San Jose 95129  
## 42 San Jose 95130  
## 43 San Jose 95131  
## 44 San Jose 95132  
## 45 San Jose 95133  
## 46 San Jose 95134  
## 47 San Jose 95135  
## 48 San Jose 95136  
## 49 San Jose 95137  
## 50 San Jose 95138  
## 51 San Jose 95139  
## 52 San Jose 95148  
## 53 San Martin 95046  
## 54 Santa Clara 95050  
## 55 Santa Clara 95051  
## 56 Santa Clara 95054  
## 57 Saratoga 95070  
## 58 Sunnyvale 94085  
## 59 Sunnyvale 94086  
## 60 Sunnyvale 94087  
## 61 Sunnyvale 94089  
## 62 Sunnyvale 94088  
## 63 Atherton 94027

While I began exploratory analysis with just several zip codes, examining the full list of zipcodes provides a more complete picture of contributions and reveals major discrepancies among the political sentiment and contribution volume of different zip codes. In addition to the insights gleaned, combining data from the individual zip codes provides a new dataset ripe for future analyses of Valley giving (most available datasets compile zipcode-level data for a specific election cycle, or a more narrow set of zipcodes).

Cleaning & Wrangling

For each contribution, the FEC data provides information on the candidate name (cand\_nm\_title), contributor name (contbr\_nm), contributor city (contbr\_city), contributor state (contbr\_st), contributor zip (contbr\_zip), contributor employer (contbr\_employer), transaction description (receipt\_desc), contribution date (contb\_receipt\_dt), and amount given (contb\_receipt\_amt). To assign the contribution to a specific party, I created a separate file listing the candidates who ran in the Democratic and Republican primaries in CA in each election cycle, denoting party affiliation. Since the names of candidates in this separate file (with tables from Wikipedia articles on the primaries) did not always match the formatting of names provided through the FEC data, I make several edits to the cand\_nm\_title variable in the main dataset to ease the merge process.

Initial analysis on several zip codes showed that in several cases, the contributions recorded were made to candidates that did not ultimately participate in the CA primary (e.g. Jeb Bush in 2016), so this list of primary participants does not capture the full set of recipients included in the dataset. Since this study does not distinguish between candidates that did/did not participate in the primary when examining party receipt totals, I manually code the party identification of those candidates missing from the primary file. The following table summarizes the candidates receiving money from Silicon Valley in each election cycle, along with a note about whether each was ultimately a candidate in the primary. N.B. Candidates (e.g. Ted Cruz) who had dropped out of the race but appeared on the CA ballot are denoted as in the primary.

## Candidate Primary?  
## 1 Alan Keyes Yes  
## 2 Barack Obama Yes  
## 3 Ben Carson Yes  
## 4 Bernie Sanders Yes  
## 5 Bill Richardson Yes  
## 6 Brian Calef (write-in) Yes  
## 7 Buddy Roemer Yes  
## 8 Christopher Dodd Yes  
## 9 David Frey (write-in) Yes  
## 10 Dennis Kucinich Yes  
## 11 Donald James Gonzales Yes  
## 12 Donald Trump Yes  
## 13 Duncan Hunter Yes  
## 14 Edward Marshall (write-in) Yes  
## 15 Eric Hinzman (write-in) Yes  
## 16 Fred Karger Yes  
## 17 Fred Thompson Yes  
## 18 H Hewes Yes  
## 19 Hillary Clinton Yes  
## 20 J Gilmore Yes  
## 21 Jeremy Hannon Yes  
## 22 Joe Biden Yes  
## 23 Joel Neuberg (write-in) Yes  
## 24 John Cox Yes  
## 25 John Edwards Yes  
## 26 John Kasich Yes  
## 27 John McCain Yes  
## 28 John Sutherland (write-in) Yes  
## 29 Joseph McAndrew (write-in) Yes  
## 30 Julius Mogyorossy (write-in) Yes  
## 31 Karen Irish (write-in) Yes  
## 32 Keith Judd (write-in) Yes  
## 33 M Steinberg Yes  
## 34 Michael Shaw (write-in) Yes  
## 35 Mike Gravel Yes  
## 36 Mike Huckabee Yes  
## 37 Mitt Romney Yes  
## 38 Newt Gingrich Yes  
## 39 Phil Epstein (write-in) Yes  
## 40 R De La Fuente Yes  
## 41 Rick Santorum Yes  
## 42 Robert Brickell (write-in) Yes  
## 43 Ron Paul Yes  
## 44 Rudy Giuliani Yes  
## 45 Sam Brownback Yes  
## 46 Sheldon Yeu Howard Yes  
## 47 Ted Cruz Yes  
## 48 Tom Tancredo Yes  
## 49 W Wilson Yes  
## 50 Walter Rothnie (write-in) Yes  
## 51 Willie Carter (write-in) Yes  
## 52 Carly Fiorina No  
## 53 Christopher J. Christie No  
## 54 Herman Cain No  
## 55 James Henry Jr. Webb No  
## 56 James R. (Rick) Perry No  
## 57 Jeb Bush No  
## 58 Jon Huntsman No  
## 59 Lawrence Lessig No  
## 60 Lindsey O. Graham No  
## 61 Marco Rubio No  
## 62 Michele Bachmann No  
## 63 Rand Paul No  
## 64 Rick Perry No  
## 65 Scott Walker No  
## 66 Timothy Pawlenty No  
## 67 Tommy G Thompson No

The FEC data shows dates in date form, so I extract year information to match years with specific election cycles. The full dataset provides the following number of observations in each year.

##   
## 2006 2007 2008 2011 2012 2014 2015 2016   
## 28 5437 27473 2863 31505 14 8720 27820

I classify contributions given in 2006, 2007, and 2008 as part of the 2008 cycle, contributions from 2011 and 2012 as part of the 2012 cycle, and contributions from 2014, 2015, and 2016 as part of the 2015 cycle. The variable year\_cycle is coded categorically to retain this information.

And initial look at the contribution amounts variable produced some puzzling observations, because the individual contributions denoted are in some cases negative. Digging deeper into the source of these negative values, I conducted background research on the description (‘receipt\_desc’) variable. The negative values relate to contribution limits (<http://www.fec.gov/pages/brochures/contrib.shtml#Presumptive_Redesignations>). Campaign treasurers must regularly check committee records to ensure successive contributions from one contributor remain within the limits. Many of the categories are ways to handle excessive contributions. While committees may deposit the excessive amount from a contribution, they must seek the contributor's *reattribution* of the portion to a joint contributor (e.g. reattribution from spouse, reattribution/redesignation requested) or the contributor's *redesignation* of the portion for a different election (redesignation from primary, redesignation to general) for which the contributor hasn't already exceeded limits. The "redesignation from" and "redesignation to” lines cancel each other out, so there is no need to remove these categories from the dataset: "REDESIGNATION TO", "REATTRIBUTION/REDESIGNATION REQUESTED", "REDESIGNATION FROM", "REDESIGNATION TO GELAC", "REDESIGNATION FROM PRIMARY", "REDESIGNATION REQUESTED", "REDESIGNATION TO GENERAL", "REDESIGNATION FROM GENERAL", "REDESIGNATION TO PRIMARY DEBT.” We keep the data on refunds ("redesignation from primary; refund to be issued") because they provide insight into the total amount given by each individual. As for the reattribution categories, "reattribution from spouse" represents net money to the candidate, and hence must be included.

The final dataset incorporates variables on a recipient’s party affiliation and the associated election cycle to the variables in the main dataframe (from FEC). The complete data frame contains the following categories:

## [1] "candidate" "contbr\_nm" "contbr\_city"   
## [4] "contbr\_st" "contbr\_zip" "contbr\_employer"   
## [7] "receipt\_desc" "month" "day"   
## [10] "year" "contb\_receipt\_amt" "date"   
## [13] "cand\_party" "year\_cycle"

**PRELIMINARY ANALYSIS**

With a cleaner dataset, I begin my analysis by testing several simple assumptions about the balance of contributions in Silicon Valley. These assumptions are based on preconceived notions of the Bay Area political spectrum, touched on briefly above. A table on the totals given to Republicans and Democrats in each election cycle enables a test of these assumptions, and also reveals several other observations.

## party cycle total\_amt  
## 1 D 2008 7921757  
## 2 R 2008 1562479  
## 3 D 2012 4953779  
## 4 R 2012 2213825  
## 5 D 2016 4192272  
## 6 R 2016 1256707

1. *Democrats received more than Republicans in 2008.*

True - Democrats received more funding ($7,921,757) than Republicans ($1,562,479) from these zipcodes in the 2008 cycle.

1. *Democrats received more than Republicans in 2012.*

True - the table above shows that this assumption also holds, with Dems receiving $4,953,779 and Reps receiving $2,213,825 - although notably the discrepency between Democratic and Republican fundraising from part is much smaller. The gap's shrinkage appears to be primarily due to a fall in funding to Democrats, in contrast to the conclusion drawn by just looking at one zipcode (as I did for preliminary analysis; data for Campbell, CA indicated that the reduced gap was due to an increase in funding to Republicans). The drop in funding to Democrats is likely because a primary was not held for the Democrats in 2012.

For a reality check on the sums returned, I looked at total contributions from Campbell, CA in the 2012 cycle. According to campaignmoney.com, Republicans received $56,882 while Democrats received $104,729. Since these values include contributions to the party committees (rather than candidates), the slightly lower values gleaned through the individual contribution data make sense.

1. *Democrats have so far received more than Republicans in 2016.*

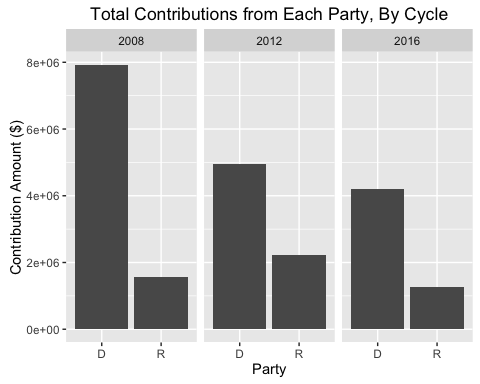
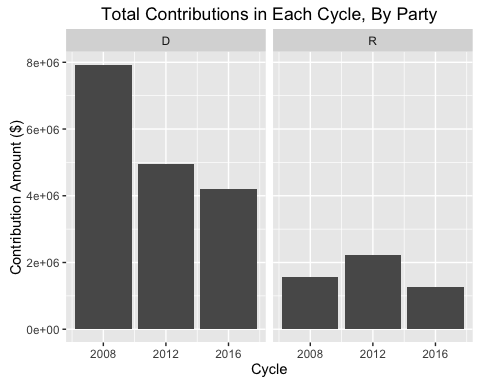
True - So far in the 2016 cycle, Democrats have so far received $4,192,272 while Republicans have received just $1,256,707.

1. *The breakdown between Democratic and Republican contributions has changed from 2008 to 2016.*

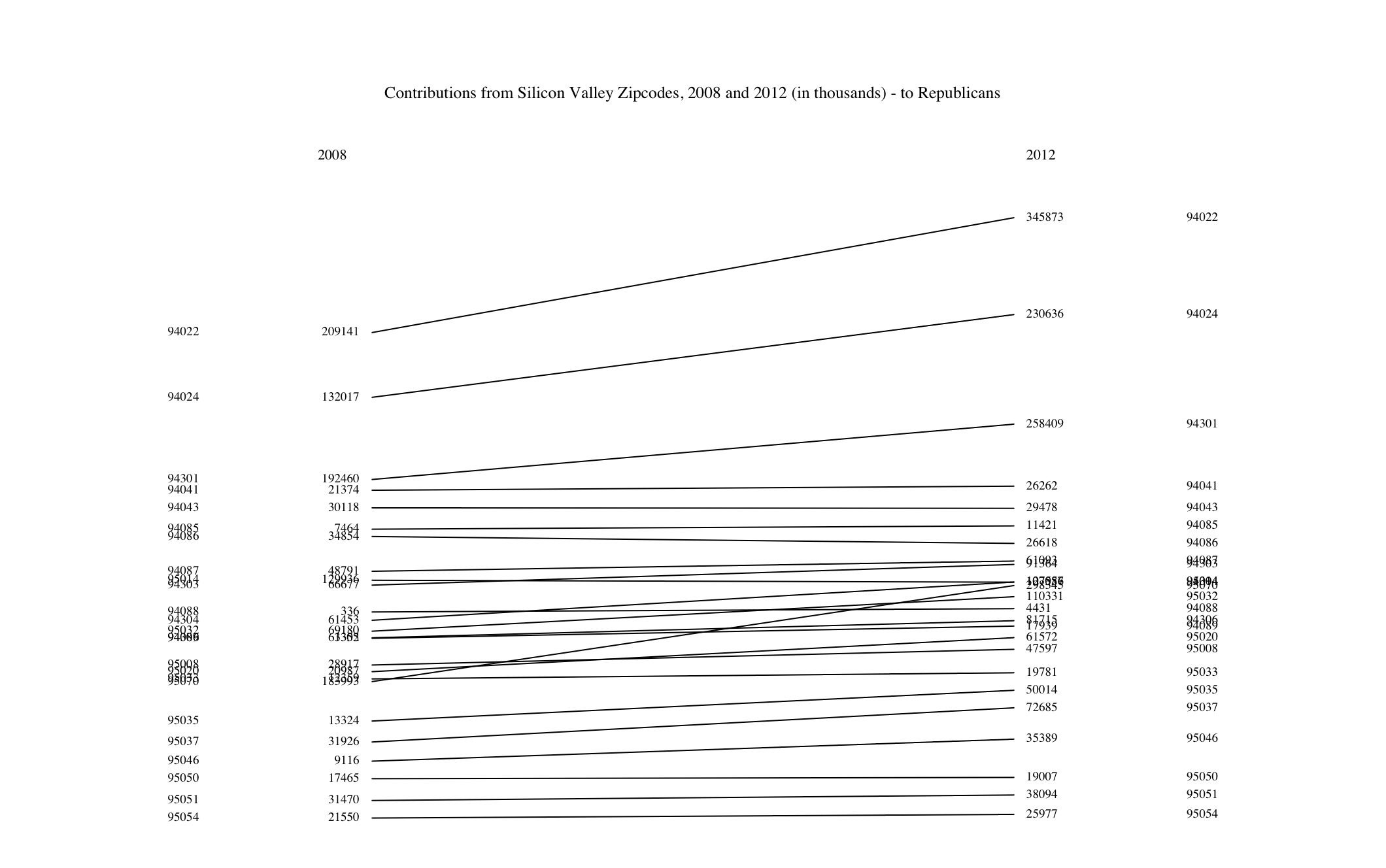
True - We can already see from the above table that the balance of fundraising shifted across the cycles, but to substantiate this numerically we can calculate the share of total funding captured by each party. This assumes that total funding is a sum of the Democratic and Republican contribution receipts, which is not the case, but the non-major party contributions can be explored further later and the relative shares will provide a basis for comparing the balance of funding between the two major parties.

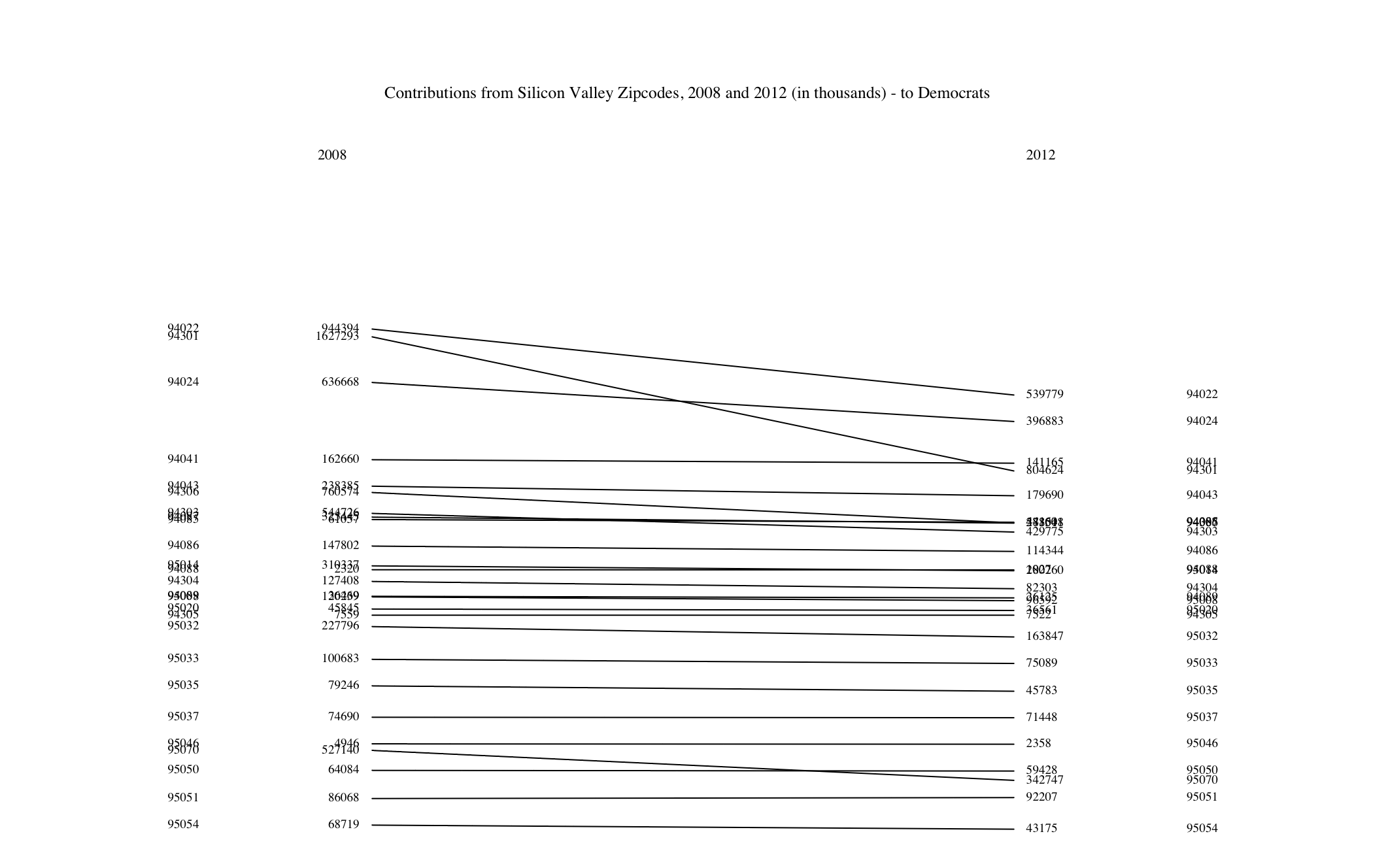
## % to Rep, 2008 % to Rep, 2012 % to Rep, 2016  
## 1 0.1647448 0.3088654 0.2306316

Incorporating additional zip codes presents similar conclusions as the test analysis for one zipcode in particular (Campbell, CA) did. There is a dramatic rise in the share of funding to Republicans in 2012, which is understandable given that the Democrats did not have a primary election in that cycle. The share to Republicans is lower in 2016 than in 2012, but 7 percentage points higher than in 2008, representing a rising share of total contributions.



To visualize trends at the zipcode level, I create a slope graph comparing total amounts from each zipcode given to a specific party in 2008 and 2012. Further analysis will explore the trends from 2008 to 2016.





**NEXT STEPS**

I will continue to explore the data to identify the best opportunities for Republican fundraising efforts in the Silicon Valley area. Specifically, my next steps will be to:

1. Calculate the bottom 10 contributing zip codes for Republicans on both an absolute ($ amount) basis and relative to the amount given to Democrats (% basis).
2. Determine which zip codes saw the biggest drops in funding between the 2016 cycle and the relevant portion of the 2008 cycle. These two cycles are the most comparable, since no incumbent is/was involved in either. However, as the 2016 cycle is not yet complete, it would be meaningless to compare funding for each party between the two cycles on an absolute basis. So I plan to compare the two on the same temporal metric through the following steps:
3. Determine the earliest contribution for the 2016 cycle
4. Calculate the days in between then and now, when the data was downloaded
5. Calculate the earliest contribution day in the data for 2008
6. Add the days from step (ii) to the value in (iii)
7. Then for the records in these time frames in 2008 and 2016, compare the sums given to Republicans - overall and by zipcode.
8. Create further visuals breaking the contribution totals down by zipcode (since the major charts so far focus on total amounts from Silicon Valley overall for each year).