

# **STATE HOUSE & SENATE ELECTION EXPENDITURE ANALYSIS**

**Akash Kadel (ak6201)**

**Jin U Bak (jub205)**

**December 16, 2016**

**Programming for Data Science**

**Master's in Data Science Program**

**New York University**

## I. INTRODUCTION:

The US House and Senate elections take place every 2 years and Federal Election Commission (FEC) provides detailed spending data for House and Senate candidates for each election period. In this project, we are interested in visualizing candidates' spending data and computing some statistics on spending behavior by candidates, states and expenditure type. We have also worked in linking this visualization to the outcome of the election, i.e. does candidate's spending behavior influence the outcome of the election?

## II. DATA ACQUISITION:

Our main sources of data<sup>1</sup> are – Candidate disbursements data and election results data for House and Senate members. The candidate disbursements contain the data related to the disbursements the candidates have made during the entire elections. The disbursements are of various types like Travel, Fundraising, Storage of Campaign records etc. In addition to this the data also contains a timestamp object that denotes the time when the expenditure was made. The results data contain the elections results for every state in US for a particular year. So the results data for three years (2010, 2012 and 2014) were pooled together with the candidate disbursements data to form a single summary table which can be used for analysis.

The election results dataset which we have is from 1982 to 2014, but the dataset for the candidate disbursements is only from 2010 to 2016. So after linking the two dataset by year, we were able to extract the information for only the years, 2010, 2012 and 2014. The linking was done by just merging the two datasets by candidate id. Candidate id is unique id assigned to each House and Senate candidate, which is constant for each candidate (even if the candidate appears in more than one elections).

### ***i. Instructions for loading data:***

*The dataset which we are using is huge. Also combining and merging will take a lot of time in the command prompt. It will generally take 1-2 minutes to load the data and perform the analysis.*

## III. CONFIGURATION

There is no other non-standard libraries to be downloaded for this program to run. It makes use of standard, built-in libraries from Python 3.X version. Please make sure you are using the right version of Python before running the program.

Once the program is downloaded or cloned from Github, you will find the following files and folders.

1. Data (Folder) – This folder contains all the necessary data that's needed for the program and analysis. All the files should be in csv format.  
Expense data files: all\_house\_senate\_2010.csv, all\_house\_senate\_2012.csv, all\_house\_senate\_2014.csv  
Result data files: result\_2010.csv, result\_2012.csv, result\_2014.csv

2. ExpenseAnalyzer (Folder) – This folder contains the following Python scripts.
  - DataHandler.py – Data Analysis tool for Election Expense data
  - interface.py – User interaction functions for the program
  - utiltools.py – Utility functions for data loading, merging and formatting.
3. main.py – This is the main script that drives the program. This can be run on command line with the following command: `python main.py`
4. test.py – This script includes unit test for the ExpenseAnalyzer module.

## IV. MAIN MENU

As soon as the user enters the command, “`python main.py`”, he is provided with the following options to perform the analysis.

```
Akash@KADEL MINGW64 ~/Desktop/ProgDS/LATEST4-20161216T110658Z/LATEST4
$ python main.py
Loading Master Data...Please wait
PROGRAM DISCRIPTION

Expense Data analysis
[1] Analysis by Year
[2] Analysis by State
[3] Year Comparison analysis
[4] Exit the program
Please enter the option number: |
```

- 
1. Expenditure by the candidates: <http://www.fec.gov/data/CandidateDisbursement.do?format=html>  
Election result: <http://www.fec.gov/pubrec/electionresults.shtml>

### i. ANALYSIS BY YEAR:

During the elections from 2010 to 2014, every year witnessed a different distribution of expenditure. The maximum amount spent in year 2010 was \$50 million and \$28 million in the year 2014. Since every election has a very different distribution of expenditure data, it is better to perform analysis for every year the election was held.

As soon as the user selects the first option from the main menu, he is asked the year he is interested in performing analysis. The following options pop up:

```
[1] Show quick summary of the year
[2] Show distribution of expenditure
[3] Show expenditures by states
[4] Show Top/Bottom 5 candidates by expenditure
[5] Show Top/Bottom N candidates by expenditure *30 Max*
[6] Show Winners VS Losers expense comparison
[7] Show monthly expenditure
[8] Change Year
[0] Main Menu
Please enter the option number: |
```

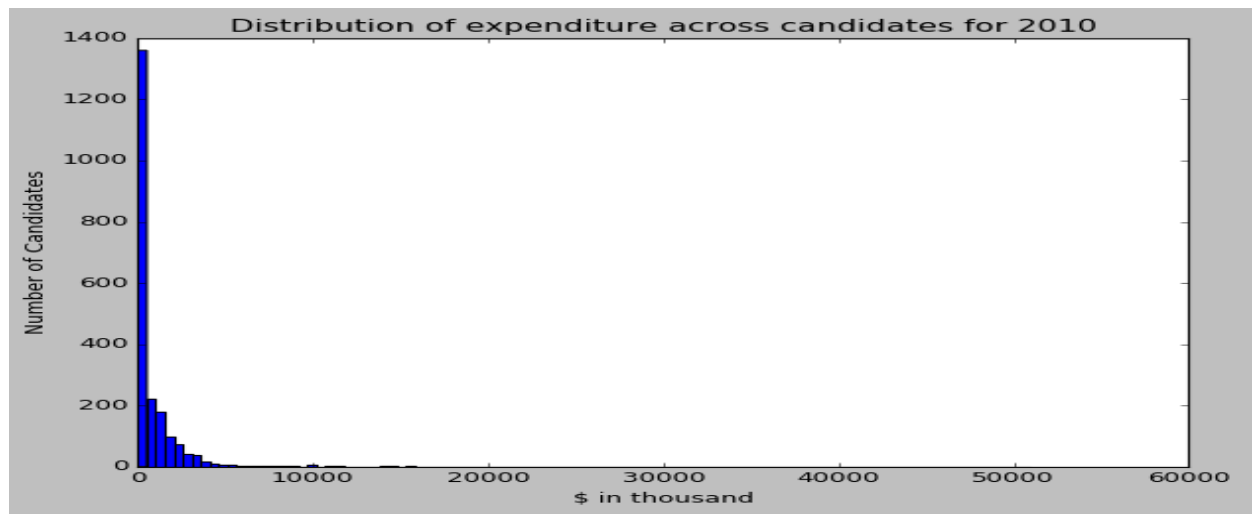
## 1. Quick Summary:

The first option is to provide a quick summary (total, min, max, median & standard deviation) of the distribution of expenditure that year. The summary statistics for the year 2010 looks like this:

```
Please enter the option number: 1
Total Number of candidates: 2107
Maximum Expense: $ 50971428.18
Minimum Expense: $ 0.10
Average Expense: $ 926435.52
Median Expense: $ 129844.10
Standard Deviation of expense: 2531020.21
```

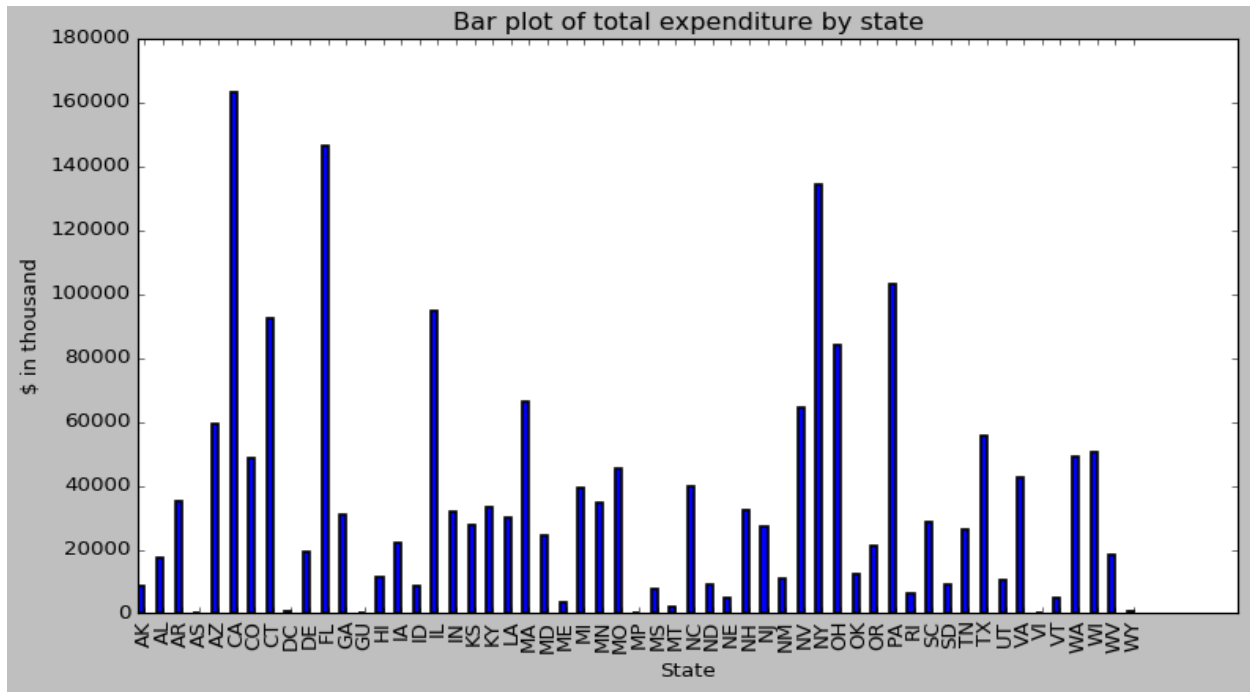
## 2. Distribution of Data:

The second option prints a graph of the distribution of the year chosen. The graph is highly skewed towards the left since there were lots of candidates who have spent less than 10 million dollars and few candidates with an expenditure of more than 20 million dollars. The maximum expense incurred by a candidate is \$50 million which can be seen from the previous statistic – Quick summary. The following graph is the distribution graph for the year 2010.



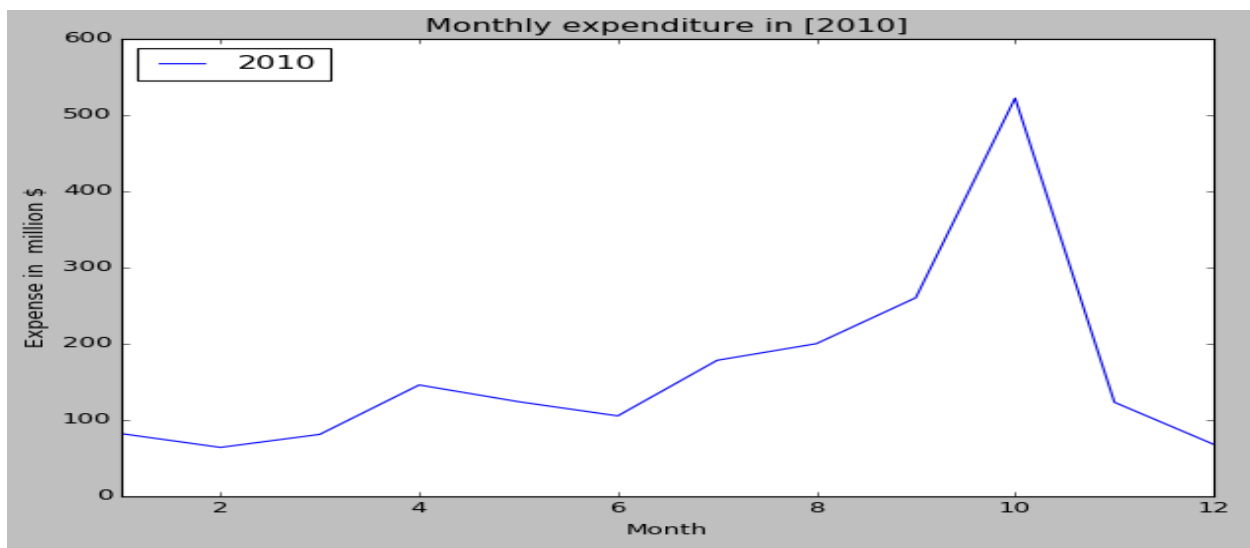
## 3. Distribution by state:

The distribution of total expenditure for a given year is different in every month of the year. This option allows us to visualize the distribution in each state. The distribution for the year 2010 look like this:



#### 4. Monthly Expenditure:

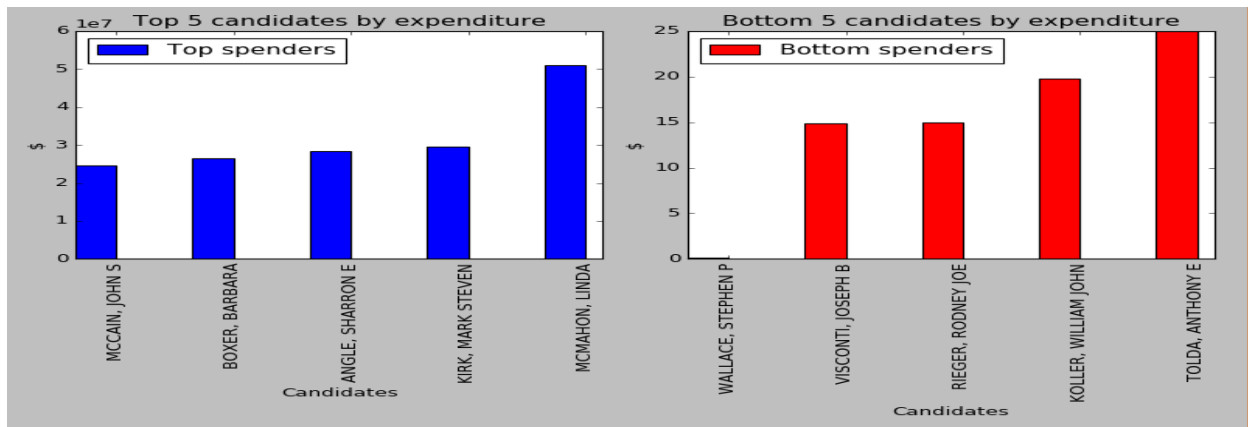
The monthly expenditure graph is basically a time series representation which shows the total amount of sales for every state for the chosen year. The monthly expenditure for the year 2010 looks like this:



From the above graph we can see that generally, candidates spend a lot during the month of October. It makes sense because the results are due during November and candidates try to maximize the outcome by investing a lot in the penultimate month of their campaign.

## 5. Top/Bottom 5 candidates by expenditure:

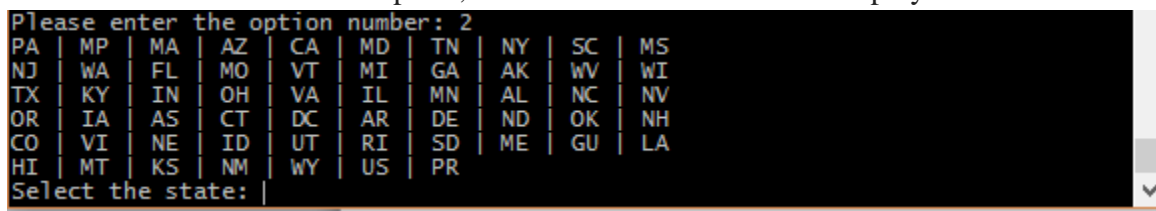
The 4th option is to see the statistical difference of the amount spent between the winners and losers. It is very useful for any candidate to have this since it helps him/her to have a basic idea of how much it is required to enable the candidate to win the elections. Here, there are two default options – to show top 5 or 10 candidates by expenditure and one custom defined input - to show top 'n' candidates by expenditure. The graph for top 5 looks:



## ii. ANALYSIS BY STATE:

The second option in the main menu is to perform the analysis for a particular state. If we go deeper into every state, we see that the distribution of the data for every state varies a lot in all the three different years. [This can be seen in the graph from the option number 3 – in analysis by year]. Also, note that there can be many winners for a single state. We can also visualize the money spent by the election winners for three different years in the same state. Infact, there are around 480 unique winners (for all 50 states collectively).

Once the user selects the 2<sup>nd</sup> option, all the state code names are displayed.

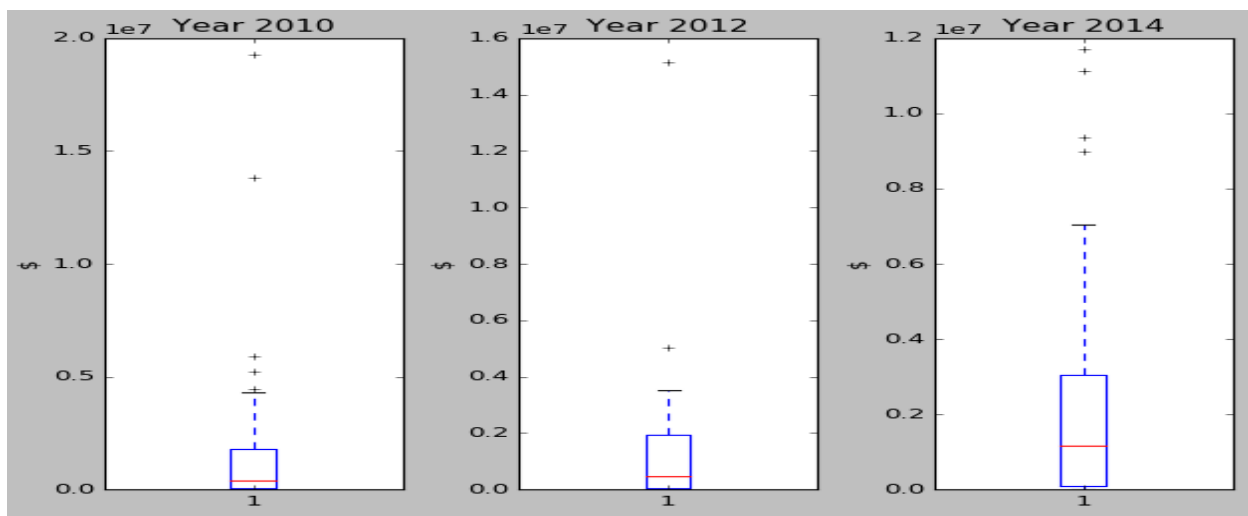


User needs to input the value as it appears in the screen. For example, for selecting New York state, user needs to enter NY (or ny). Also note that the analysis is performed across all the years for the given state

```
Select the state: NY
[1] Show quick summary of the state
[2] Show distribution of expenditure
[3] Show monthly expenditure
[4] Show winner vs loser comparison
[8] Change State
[0] Main Menu
Please enter the option number: :
```

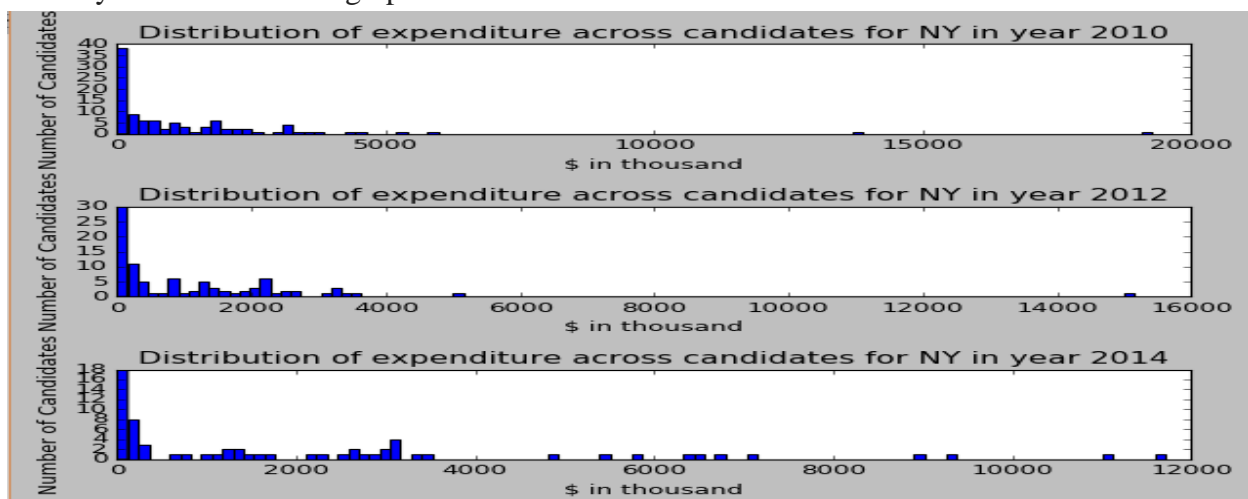
## 1. Quick Summary of the state:

The quick summary option displays the basic statistics for the chosen month. We can see the distribution highly skewed in 2010 whereas in 2014, the distribution is almost equally spread out between 1 million and 10 million.



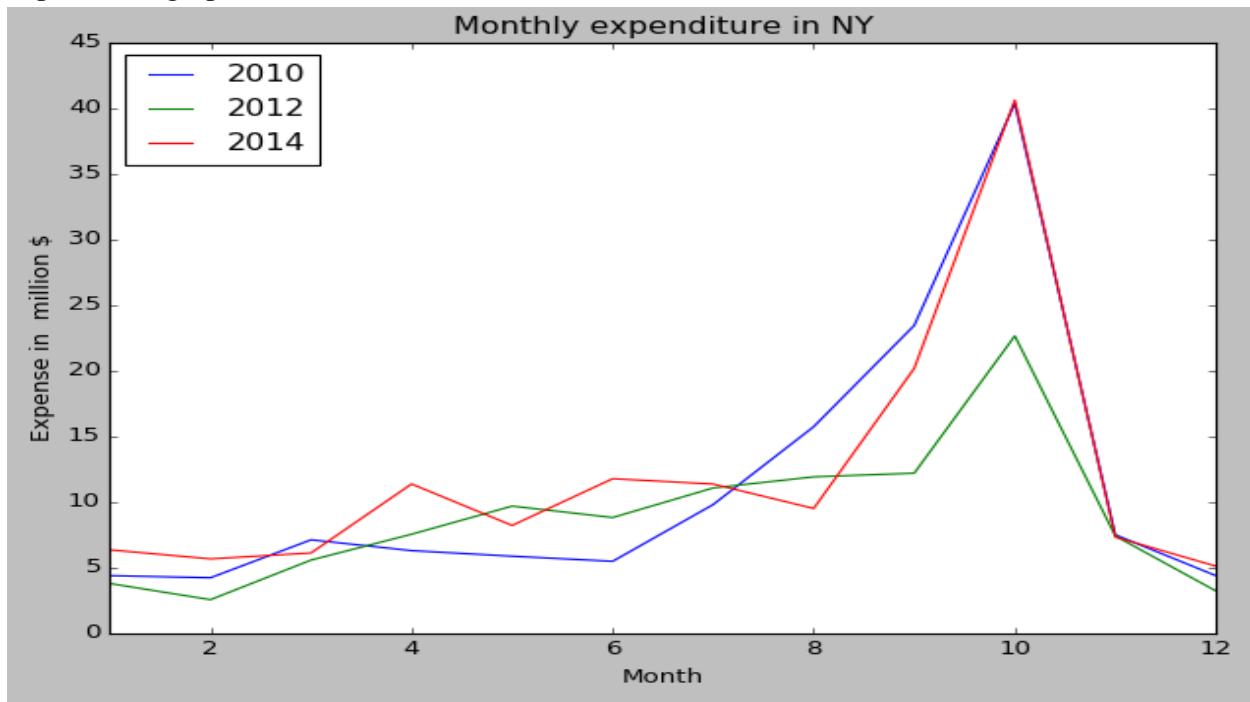
## 2. Distribution of expenditure in different years:

The second option displays the distribution of expenditure for the years 2010, 2012 and 2014 in separate graphs. For the state New York, we can see that the distribution is varying a lot in every election from the graph below.



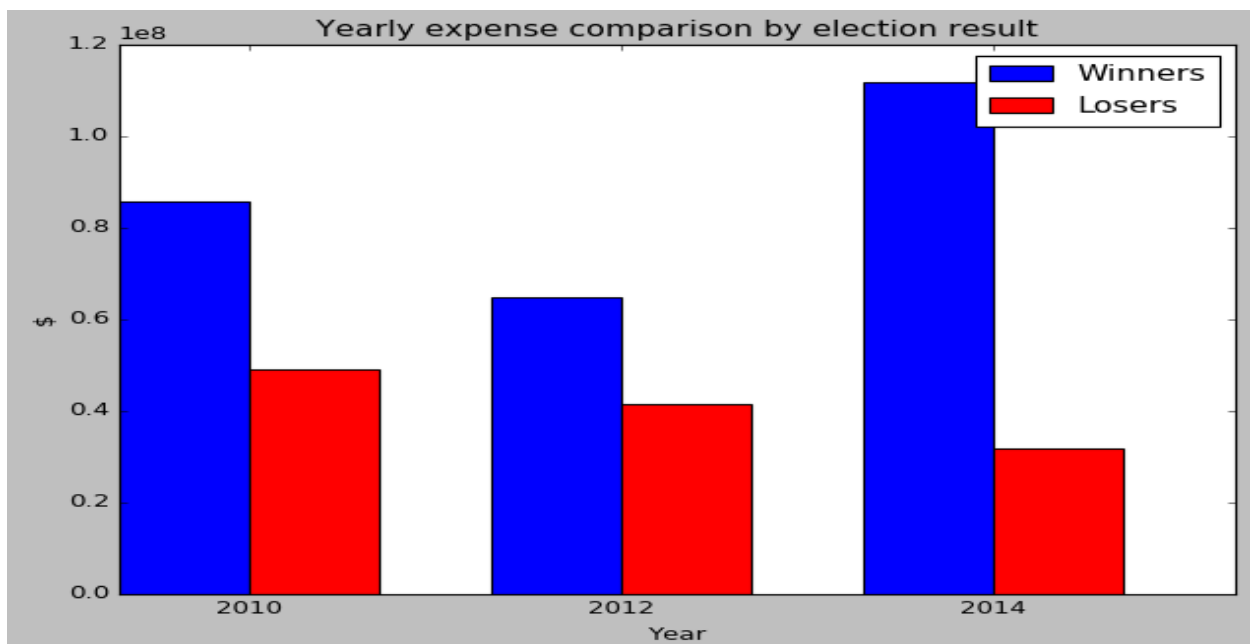
### 3. Monthly expenditure:

Monthly expenditure is like a time series graph for all the three years. It is used to visualize the distribution of expenditure across different months of the year. The monthly expenditure graph for New York stats looks like this.



### 4. Winner vs Losers comparison:

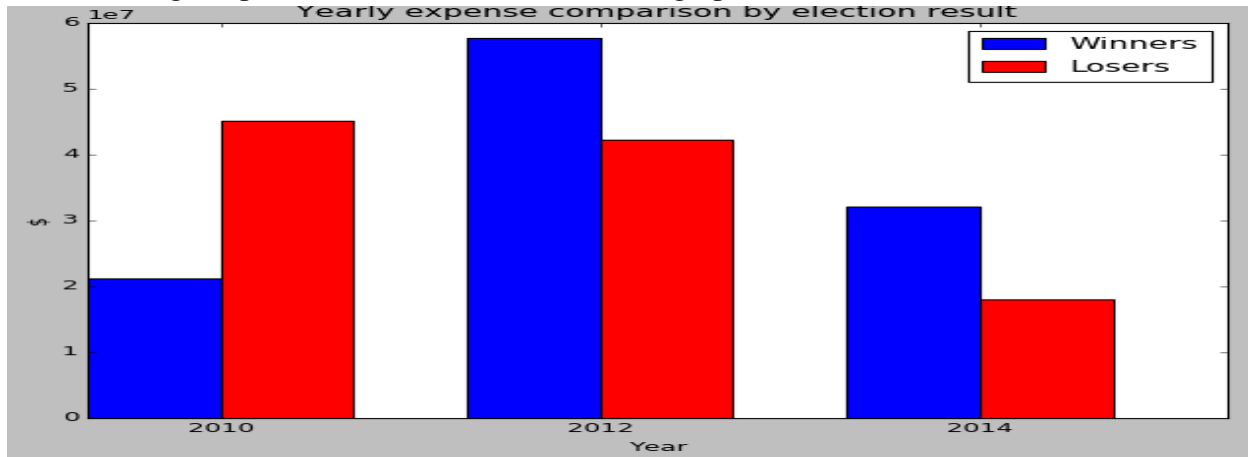
Winner vs loser comparison graph plots the total amount spent by the winners vs the total amount spent by the losers in the different years. It gives us the intuition of expenditure a candidate has to incur in campaigning to have a good chance in the elections. The graph for New York State looks like:





From the graph we can see the difference in amount spent by winners and losers is very high. Winners here spend a lot compared to the losers through campaigning, to have a lead in the elections. Whereas losers didn't incur that much expenditure as winners did.

But in some cases we can also see that losers are spending more than winners. This might be because some candidates are guaranteed winning because of their popularity and preferred choice among the public. For instance, consider the graph for MASSACHUSETTS.



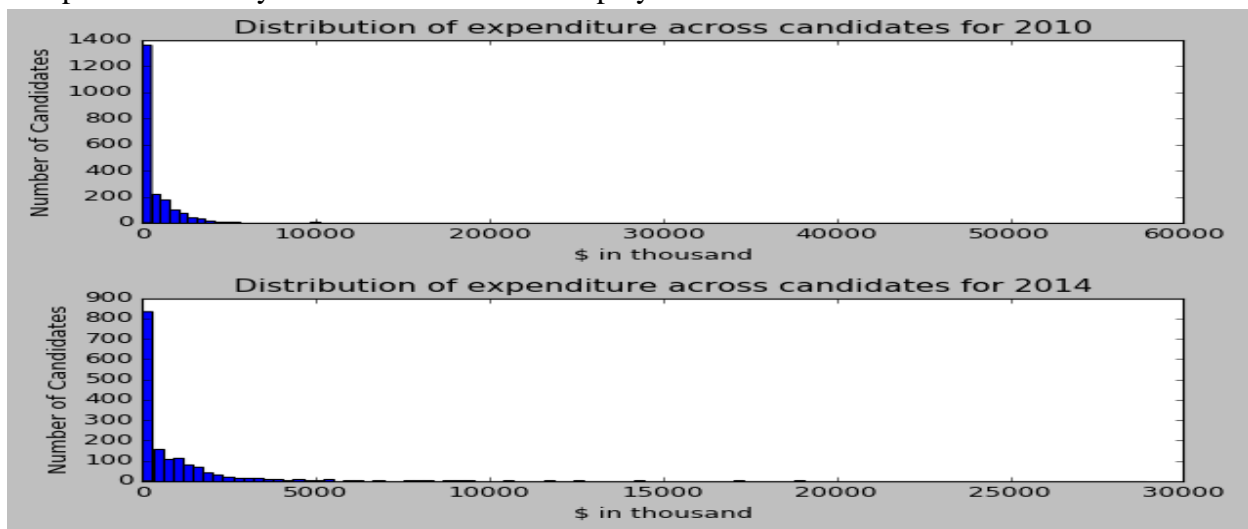
### iii. YEAR WISE COMPARISON:

This option is to see how much the distribution is varying during every election year. To start the comparison, the user needs to enter the values in comma separated format.

```
Expense Data analysis
[1] Analysis by Year
[2] Analysis by State
[3] Year Comparison analysis
[4] Exit the program
Please enter the option number: 3
Please enter the years you are interested in(comma separated): 2010, 2012
```

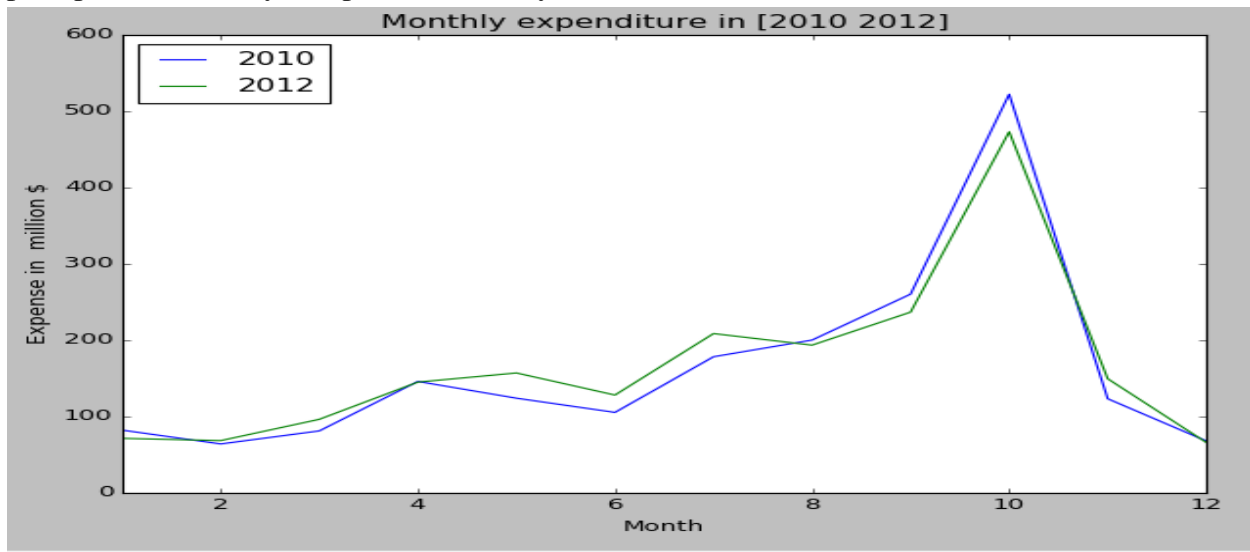
#### 1. Distribution by years:

This is to compare the distribution of expenditures for all the year. The year wise comparison for the years 2010 and 2012 is displayed below.



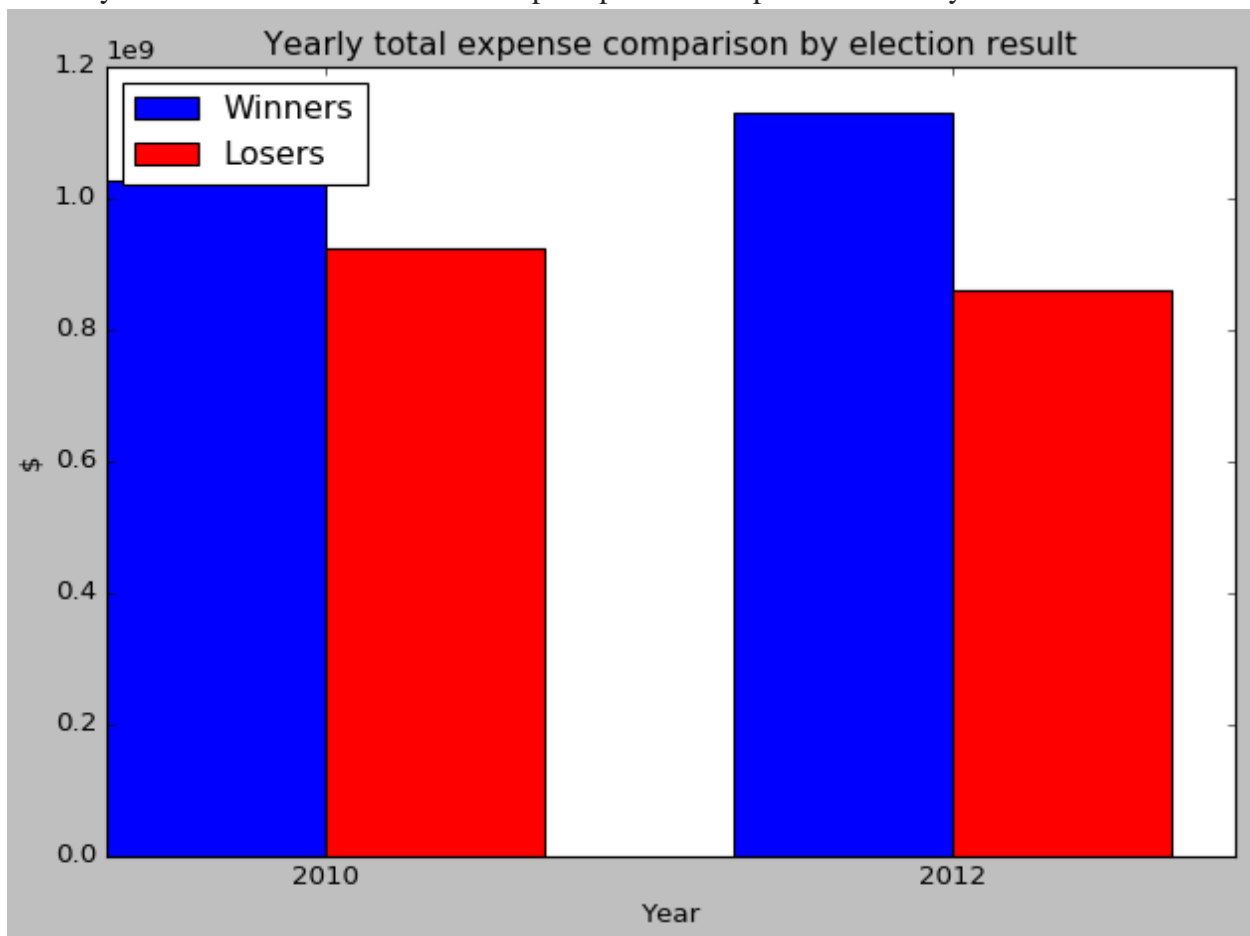
## 2. Show monthly Expenditure:

This is to make the monthly wise comparison for the years mentioned in the command prompt. The monthly comparison for the years 2010 and 2012 is :



## 3. Winner vs Loser comparison:

This option is to show the comparison of total amount incurred by the winners and losers for the years mentioned in the command prompt. The comparison for the years 2010 and 2012 is



## V. DEMOS TO RUN A QUERY:

Suppose a user wants to accumulate the following information:

- (i) To see the total expenditure in the year 2010
- (ii) To see the distribution across all the states in USA
- (iii) To see top and bottom 5 candidates by expenditure
- (iv) To see the expenditure distribution for Arizona for all the years – 2010,2012,2014
- (v) To see the monthly distribution for Arizona for all the years - 2010,2012,2014

### (i) To see the total expenditure in the year 2010:

Step 1: Enter “python main.py”

```
Akash@KADEL MINGW64 ~/Desktop/ProgDS/LATEST4
$ python main.py
Loading Master Data...Please wait
PROGRAM DISCRIPTION
```

Step 2: Choose “Analysis by Year” by entering 1

```
Expense Data analysis
[1] Analysis by Year
[2] Analysis by State
[3] Year Comparison analysis
[4] Exit the program
Please enter the option number: 1
```

Step 3: Enter the year 2010

```
Expense Data analysis
[1] Analysis by Year
[2] Analysis by State
[3] Year Comparison analysis
[4] Exit the program
Please enter the option number: 1
Please enter the year you are interested in: 2010
```

Step 4: Out of all the options choose the first option “Show quick summary of the year” by entering 1

```
[1] Show quick summary of the year
[2] Show distribution of expenditure
[3] Show expenditures by states
[4] Show Top/Bottom 5 candidates by expenditure
[5] Show Top/Bottom N candidates by expenditure *30 Max*
[6] Show Winners VS Losers expense comparison
[7] Show monthly expenditure
[8] Change Year
[0] Main Menu
```

Step 5: See the summary and total expenditure

**(ii) To see the distribution across all the states in USA:**

Step 1: After returning back to the menu for Analysis by Year, select “distribution of the expenditure across all states” by entering 3

```
[1] Show quick summary of the year
[2] Show distribution of expenditure
[3] Show expenditures by states
[4] Show Top/Bottom 5 candidates by expenditure
[5] Show Top/Bottom N candidates by expenditure #30 Max*
[6] Show Winners VS Losers expense comparison
[7] Show monthly expenditure
[8] Change Year
[0] Main Menu
```

Step 2: You will see a bar graph with all the states in the x axis, and the expenditure in the y axis.

**Note: PLEASE CLOSE THE FIGURE TO CONTINUE THE ANALYSIS**

**(iii) To see top and bottom 5 candidates by expenditure**

Step 1: After returning back to the menu for Analysis by Year, select “Show Top/Bottom 5 candidates by expenditure” by entering 3

```
[1] Show quick summary of the year
[2] Show distribution of expenditure
[3] Show expenditures by states
[4] Show Top/Bottom 5 candidates by expenditure
[5] Show Top/Bottom N candidates by expenditure #30 Max*
[6] Show Winners VS Losers expense comparison
[7] Show monthly expenditure
[8] Change Year
[0] Main Menu
```

**(iv) To see the expenditure distribution for Arizona for all the years – 2010,2012,2014:**

Step 1: Return to the MAIN MENU by hitting ‘0’ in the keyboard. The main menu will appear.

Step 2: Select the option, “Analysis by State” by entering 2 in the keyboard:

```
Expense Data analysis
[1] Analysis by Year
[2] Analysis by State
[3] Year Comparison analysis
[4] Exit the program
Please enter the option number: 2|
```

Step 3: You will get to see all the states by their short form. Enter 'AZ' to perform the analysis for Arizona:

```
Please enter the option number: 2
PA | MP | MA | AZ | CA | MD | TN | NY | SC | MS
NJ | WA | FL | MO | VT | MI | GA | AK | WV | WI
TX | KY | IN | OH | VA | IL | MN | AL | NC | NV
OR | IA | AS | CT | DC | AR | DE | ND | OK | NH
CO | VI | NE | ID | UT | RI | SD | ME | GU | LA
HI | MT | KS | NM | WY | US | PR
Select the state: AZ
```

Step 4: Select the second option, "Show distribution of expenditure" by entering 2

```
Select the state: AZ
[1] Show quick summary of the state
[2] Show distribution of expenditure
[3] Show monthly expenditure
[4] Show winner vs loser comparison
[5] Change State
[0] Main Menu
Please enter the option number: 2
```

**(v) To see the monthly distribution for Arizona for all the years - 2010,2012,2014**

Step 1: After returning back to the menu for ANALYSIS BY STATE, choose the 3<sup>rd</sup> option by entering 3 in the command prompt.

```
[1] Show quick summary of the state
[2] Show distribution of expenditure
[3] Show monthly expenditure
[4] Show winner vs loser comparison
[5] Change State
[0] Main Menu
Please enter the option number: 3
```

Step 2: A graph will pop up which will show monthly distribution of expenditure for ARIZONA for all the years.

