IT497 OSMEN Assignment

Can Facebook Predict the Illinois Governor's Race?

Author - Shweta Vyas

Introduction

The Illinois State elections, to be held on November 4,2014, in which incumbent Democratic Governor Pat Quinn and Republican candidate Bruce Rauner are contesting for the position of Governor of Illinois. Rauner's primary campaign agenda for this election is to *bring jobs to Illinois State*. While Pat Quinn's primary agenda is to *improve ethical standards and protect public sector labor unions*. In recent times Social media has become quite influential in bringing to light all the political happenings in any part of the world. Politicians these days put all their campaign details and spend huge amount of money in IT to increase the likes, shares and followers count. Social media is not only beneficial to the candidates but also to the voters who can closely follow the activities and views of each candidate and even share their opinion - thereby making a sensible decision of who deserves their vote in the upcoming election. Through this Election analysis report, one can clearly see the popularity of Bruce Rauner in the upcoming Illinois Governor's Election. The number of people following and sharing the posts of Bruce Rauner in year 2014 are much more as compared to Pat Quinn. This is contradictory to the amount of money spent by Republicans and Democratics in Social media.

1. O - Obtaining the Data

In the below code chunk, commands from three R libraries i.e. rjson, httr and Rfacebook are used. In order to fetch the facebook data one first need to generate a "token". The token can be temporary or permanent. In this paper, temporary token is generated by clicking on the link Graph API Explorer and using a valid Facebook login credentials.

```
library(rjson);
library(httr);
library(Rfacebook);
```

```
token<-"CAACEdEose0cBANBuy1rReYQ3CHAGaxJENSevUjXD5nwHHCrm8m6cYZBtFzYzPXUE52HH0BJVLq5LeZA7ZAlZCLKDZCVSyvj7on4
iJlVhNbbIHPOvdgscdOa16sE3V1fAAQjJjGbhiOZC6i7kTqA7WHsfZAIXV4sdBgdkHeENEwoVIi1h7jQZBZCU2sTTnjItqSsKIF4Ex4vzCUr
oQcjzUcZAdW";
```

Once the token is created, we can use getPage commands from Rfacebook package to fetch the number of likes, shares and posts of candidates along with the posted date and time.

```
page_BruceR<-getPage("BruceRauner",token, n = 100000);
```

```
## 100 posts 200 posts 300 posts 400 posts 500 posts 570 posts
```

```
page_PatQ<-getPage("GovernorQuinn",token, n = 100000);
```

```
## 100 posts 200 posts 300 posts 400 posts 500 posts 600 posts 700 posts 800 posts 886 posts 986 posts 1086 posts 1186 posts 1286
posts 1354 posts
```

2. S - Scrubbing the Data

Here Bruce Rauner Facebook page and its posts are taken into the dataframe called page_BruceR. Then this data is cleaned in such a way that records from current year, “2014” are taken into consideration - as the candidates start their campaigning since the beginning of this year. To achieve this, date format in created_time column is changed using *format(as.date(column_name))* command to DD-MM-YYYY format and then searching the pattern “2014” in the created time column through *grep* command.The *format* is a command in the base package through which the date format is changed from standard %Y%M%D format to any desired format. The *grep* command is used to get the observations with the matching argument pattern(in our case pattern is 2014).

After this, I have considered only the columns essential for counting the shares ie from_id, from_name, created_time, type and shares_count. Finally I have used sum command to add the values in the shares_count column of the dataframe which gives **the total number of times Bruce Rauner’s posts are shared by his followers in the year 2014.**

```
BruceR<-page_BruceR[order(page_BruceR$type),];
BruceR$created_time<- format(as.Date(BruceR$created_time),"%d/%m/%Y");
BruceRauner<-BruceR[rev(order(BruceR$created_time)),];
BruceRauner.clean1<-BruceRauner[c(1,2,4,5,10)];
BruceRauner.clean2<-BruceRauner.clean1[grep("2014",BruceRauner.clean1$created_time),];
sumShareBruce<-sum(BruceRauner.clean2$shares_count);
```

Here Pat Quinn’s Facebook page and its posts are taken into a dataframe named page_PatQ. Deploying similar cleaning techniques for Pat Quinn as used for Bruce Rauner, we achieve the final dataframe for Pat Quinn’s facebook page. Finally adding the values in the shares_count column of the dataframe, gives us the total number of times Pat Quinn’s posts are shared by his followers in the year 2014.

```
page_PatQ$created_time<- format(as.Date(page_PatQ$created_time),"%d/%m/%Y");
PatQuinn.clean1<-page_PatQ[c(1,2,4,5,10)];
PatQuinn.clean2<-PatQuinn.clean1[grep("2014",PatQuinn.clean1$created_time),];
sumSharePat<-sum(PatQuinn.clean2$shares_count);
```

3. M - Modeling the Data

After getting the total number of shares of both the candidates, a new dataframe is created which contains the name of candidates and their corresponding total shares' count. To create a data frame, *data.frame* command is used.

```
Candidate <- c("Pat Quinn","Bruce Rauner");
Share_Count<- c(sumSharePat,sumShareBruce);
NextGov<-data.frame(Candidate,Share_Count);
```

Final dataframe, **NextGov** is obtained and the popularity of the both the candidates is clearly analysed by observing the number of times the posts of each candidate is shared. To view the final data frame we can simply write the data frame name.

```
NextGov;
```

```
##      Candidate Share_Count
## 1    Pat Quinn        4797
## 2 Bruce Rauner       14997
```

Table 1: NextGov Dataframe showing Total Number of posts shared for Bruce Rauner and Pat Quinn

4. E - Exploring :- Data Section

This section shows the class of the dataframe and its columns. The class shows the data tpye pf variables present in the columns when applied to the data frame. The structure of the dataframe is obtained through the *str* command. Summary command gives the mean, minimum , maximum, median of the dataframe.

```
class(NextGov);
```

```
## [1] "data.frame"
```

```
class(Candidate);
```

```
## [1] "character"
```

```
class(Share_Count);
```

```
## [1] "numeric"
```

```
class(NextGov$Candidate);
```

```
## [1] "factor"
```

```
class(NextGov$Share_Count);
```

```
## [1] "numeric"
```

```
str(NextGov);
```

```
## 'data.frame':    2 obs. of  2 variables:
##  $ Candidate  : Factor w/ 2 levels "Bruce Rauner",..: 2 1
##  $ Share_Count: num  4797 14997
```

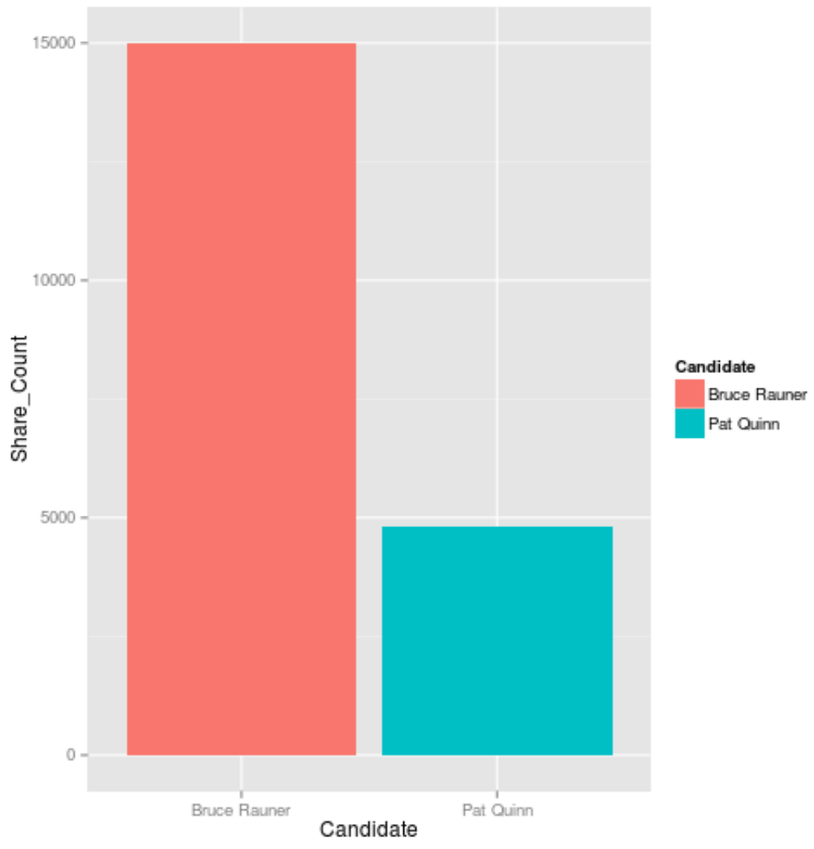
```
summary(NextGov);
```

```
##         Candidate  Share_Count
##  Bruce Rauner:1   Min.   : 4797
##  Pat Quinn   :1   1st Qu.: 7347
##                   Median : 9897
##                   Mean   : 9897
##                   3rd Qu.:12447
##                   Max.   :14997
```

5. N - GraphiNg Section

Below is the bar graph showing the count of posts shared for both the candidates. As seen through the graph, Bruce Rauner posts' shares_count is more than that of Pat Quinn. Bruce Rauner popularity in the social media is quite evident through this graph. One reason for this might be the Tax increase under Pat Quinn's governance which has raised an anguish among the citizens of Illinois State and hence increasing the followership of opponent Bruce Rauner. Bruce Rauner step farther ahead by promising People of Illinois state to bring jobs to the state. Hence the graph shows more shares count for Bruce Rauner. ggplot command is present in library named ggplot. The *data* attibute of the ggpot contains the name of data frame, *aes>/i> contains the information about the x axis and yaxis and geom_bar function creates a bar graph with stat as "identity" which shows the height of the bar graph as the value in the column.*

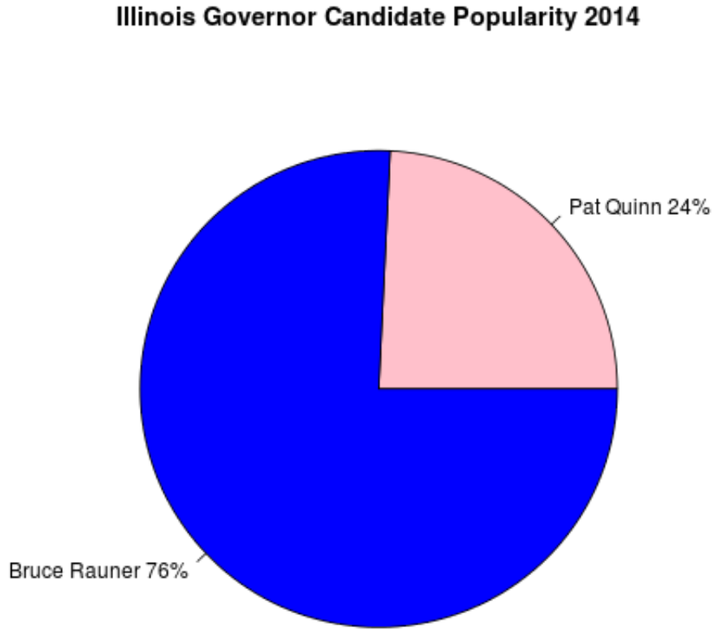
```
library(ggplot2);
ggplot(data=NextGov, aes(x=Candidate, y=Share_Count, fill=Candidate)) + geom_bar(stat="identity");
```



Plot 1: Total Number of posts shared for Bruce Rauner and Pat Quinn.

Usually result of elections are given based on the percentage of people who voted and then percentage of votes received by each candidate. So when we calculate the percentage of shares in year 2014, Bruce Rauner shows higher percentage than incumbent Governor Pat Quinn. Here round function rounds off the result of the expression inside it. To add the % symbol and labels to the pie chart I have used paste command.Finally using the pie command(present in the base package) I have created a pie chart. Through col attribute color is given and through attribute main we can give the title of the pie chart.

```
#Pie Chart With Percentages
percent_popularity<-round(NextGov$Share_Count/sum(NextGov$Share_Count)*100);
NextGov$Candidate<-paste(NextGov$Candidate,percent_popularity);
NextGov$Candidate<-paste(NextGov$Candidate,"%",sep="");
pie(NextGov$Share_Count,labels=NextGov$Candidate,col=c("pink","blue"),main="Illinois Governor Candidate Popularity 2014");
```



Plot 2: Percentage of posts shared for Bruce Rauner and Pat Quinn

Conclusion

Social media is quite beneficial in predicting and analysing elections these days but at the end of the day there are several other factors that need to be taken into account when analysing the winner of any upcoming elections. These include - people who vote in the elections do they follow chosen candidate on Social media also, changing trends in followership of several social media sites, illegally increase of likes and share, and many more. Nonetheless, Social media has become a key strategical domain to highlight the every minute political fervor to general public. Conclusively as per the extracted facebook report Bruce Rauner is leading as large number of people share his posts but the actual picture will unveils on the D-day, the Election Day.

References:

Bruce Rauner Ballatopedia

Pat Quinn Ballatopedia

Pat Quinn Facebook Page

Bruce Rauner Facebook Page

CRAN R Tutorial