

# IT497 OSEMN Assignment

## Can Facebook predict Illinois Governor's Race

### Introduction

The purpose of this paper is to understand and analyze whether Facebook.com –which is the most popular social networking site- can be used to make predictions about Illinois Governor's Race. Facebook is a social networking site which is loved by millions of people across the globe. Using Facebook, people can share their views and thoughts with each other's. People can create their pages, profiles and discussion forums on the Facebook. So on this platform, popularity of famous personalities can be judged by viewing the number of likes, followers and comments on their pages.

Using language R, this report predicts the result of election for the Illinois governor's position by fetching the data from Facebook. Pat Quinn and Bruce Rauner are two candidates who are competing in the election of Nov, 2014 for the position of the governor of Illinois. I have taken data about their followers from their respective Facebook pages to make the predictions about the most probable winner of the election based on the popularity. In this analysis, Rfacebook is the package which is used to apply various functions to get number of likes on photos, videos and comments on Facebook page of each candidate.

### Obtaining the data

In this section the process of obtaining the data from Facebook is explained. To obtain data from the facebook, first a person needs to have an account on [www.facebook.com](http://www.facebook.com). Then, we need temporary access token from <https://developers.facebook.com/tools/explorer> and this token expires in 2 hours. So, in every 2 hours we need to take new token to carry out our analysis. Below is the token I have taken to carry out the analysis. In the below code "token" is the variable which stores the value of current access token acquired from the facebook.com

```
token="CAACEdEose0cBAJYUZAKIgvRBz2GD4dp03c08zrKBb02VHZBKmbYDb36UsGLZATLZBDCZB3euZAnG8v8hFtELRCawadJgzS63QZCM1C  
NrAuNZBuLiZC2ZCiAPdrzBtdCZAhd8nQJCOArsICiSe5wT7ph9VfmgYF9tz17mZB7UMbHLzSSZAGFZBP7p1627m55IgZBHZChk43VG973BuIZ  
BJQTO8YSN1XB3t"
```

In the below code chunk, required libraries to carry-out our analysis are given. Rfacebook is a package which provides us a various commands to find number of likes, shares, friends and followers of pages and people available on Facebook. To get use of commands provided by Rfacebook package, libraries httr, rjson and Rfacebook must be available in our code.

1. httr
2. rjson
3. Rfacebook

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

In the below code, we are saving all the information like Page Id, status, photos count, likes count, videos count etc. from Pat Quinn's facebook page into the data frame named page\_GovernorQuinn. getPage Command brings all the data from the requested facebook page into the data frame.

I have provided the value of access token in the token variable. "n" determines limit to fetch records from the Facebook page. I have given very large value to "n" so that all the records can get fetched from the facebook page.

Information available on the facebook page of Pat Quinn. The result of below command is the total number of post available on Pat Quinn’s facebook page. (100 posts are fetched per request/response cycle)

```
page_GovernorQuinn <- getPage("GovernorQuinn", token=token, n=10000000000);
```

```
## 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
```

Information available on the facebook page of Bruce Rauner. The result of below command is the total number of post available on Bruce Rauner’s facebook page. (100 posts are fetched per request/response cycle)

```
page_BruceRauner <- getPage("BruceRauner", token=token, n=10000000000);
```

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

**Data Scrubbing**

In this section, scrubbing of data collected from facebook is explained. Data is of no use if it’s not clean enough to get/create some useful information. Using below code we will get total number of likes on status, photos and videos on the facebook page of each competitor.

1. Data scrubbing for data gathered from Pat Quinn’s Facebook page

In the below code, page\_GovernorQuinn\$type is the column naming “type” of data frame page\_GovernorQuinn which has information about type of post like photo, video, status etc. grep command on page\_GovernorQuinn\$type column with “photo” provides the list of photos in this column. Lastly, sum(photo\_GovernorQuinn\$likes\_count) gives us total number of likes on all the photos of Pat Quinn on his facebook page.

```
photo_GovernorQuinn=page_GovernorQuinn[grepl("photo",page_GovernorQuinn$type),];
photo_GovernorQuinn_likes=sum(photo_GovernorQuinn$likes_count);
```

In the below code, page\_GovernorQuinn\$type is the column naming “type” of data frame page\_GovernorQuinn which has information about type of post like photo, video, status etc. grep command on page\_GovernorQuinn\$type column with “status” provides the list of status in this column. Lastly, sum(status\_GovernorQuinn\$likes\_count) gives us total number of likes on all the status of Pat Quinn on his facebook page.

```
status_GovernorQuinn=page_GovernorQuinn[grepl("status",page_GovernorQuinn$type),];
status_GovernorQuinn_likes=sum(status_GovernorQuinn$likes_count);
```

In the below code, page\_GovernorQuinn\$type is the column naming “type” of data frame page\_GovernorQuinn which has information about type of post like photo, video, status etc. grep command on page\_GovernorQuinn\$type column with “video” provides the list of video in this column. Lastly, sum(video\_GovernorQuinn\$likes\_count) gives us total number of likes on all the video of Pat Quinn on his facebook page.

```
video_GovernorQuinn=page_GovernorQuinn[grepl("video",page_GovernorQuinn$type),];
video_GovernorQuinn_likes=sum(video_GovernorQuinn$likes_count);
```

2. Data scrubbing for data gathered from Bruce Rauner’s Facebook page

In the below code, `page_BruceRauner$type` is the column naming “type” of data frame `page_BruceRauner` which has information about type of post like photo, video, status etc. `grep` command on `page_BruceRauner$type` column with “photo” provides the list of photos in this column. Lastly, `sum(photo_BruceRauner$likes_count)` gives us total number of likes on all the photos of Bruce Rauner on his facebook page.

```
photo_BruceRauner=page_BruceRauner[grep("photo",page_BruceRauner$type),];
photo_BruceRauner_likes=sum(photo_BruceRauner$likes_count);
```

In the below code, `page_BruceRauner$type` is the column naming “type” of data frame `page_BruceRauner` which has information about type of post like photo, video, status etc. `grep` command on `page_BruceRauner$type` column with “status” provides the list of status in this column. Lastly, `sum(status_BruceRauner$likes_count)` gives us total number of likes on all the status of Bruce Rauner on his facebook page.

```
status_BruceRauner=page_BruceRauner[grep("status",page_BruceRauner$type),];
status_BruceRauner_likes=sum(status_BruceRauner$likes_count);
```

In the below code, `page_BruceRauner$type` is the column naming “type” of data frame `page_BruceRauner` which has information about type of post like photo, video, status etc. `grep` command on `page_BruceRauner$type` column with “video” provides the list of video in this column. Lastly, `sum(video_BruceRauner$likes_count)` gives us total number of likes on all the video of Bruce Rauner on his facebook page.

```
video_BruceRauner=page_BruceRauner[grep("video",page_BruceRauner$type),];
video_BruceRauner_likes=sum(video_BruceRauner$likes_count);
```

After collecting the data about the number of likes on status, photos and videos of each candidate, a data frame to store the final results(numbers) is created. The data frame is stored in the variable naming `final.df`. Below is the code to create data frame. In variable “Name”, Pat Quinn and Bruce Rauner is saved. In the variable `Photo_likes`, number of likes on photos of Pat Quinn and Bruce Rauner is saved. In the variable `Video_likes`, number of likes on videos of Pat Quinn and Bruce Rauner is saved. In the variable `Status_likes`, number of likes on status of Pat Quinn and Bruce Rauner is saved.

```
final.df=data.frame(Name=c("Pat Quinn","Bruce Rauner"),Photo_likes=c(photo_GovernorQuinn_likes,photo_BruceRauner_likes),Video_likes=c(video_GovernorQuinn_likes, video_BruceRauner_likes), Status_likes=c(status_GovernorQuinn_likes,status_BruceRauner_likes))
```

Below is the data frame holding final numbers about popularity of Pat Quinn and Bruce Rauner. That means, this data frame holds the data about number of likes on photos, status and videos on facebook page of Pat Quinn and Bruce Rauner.

final.df

##	Name	Photo_likes	video_likes	Status_likes
## 1	Pat Quinn	28815	1082	13175
## 2	Bruce Rauner	73623	20045	37377

Table 1 shows the cleaned data collected from the data sources.

Data Section

1. Datatype of final.df

```
class(final.df)
```

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

2. Datatype of page\_GovernorQuinn

```
class(page_GovernorQuinn)
```

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

3. Datatype of page\_BruceRauner

```
class(page_BruceRauner)
```

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

4. Datatype of photo\_BruceRauner

```
class(photo_BruceRauner)
```

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

5. Datatype of photo\_GovernorQuinn

```
class(photo_GovernorQuinn)
```

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

6. Datatype of photo\_BruceRauner\_likes

```
class(photo_BruceRauner_likes)
```

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

7. Datatype of photo\_GovernorQuinn\_likes

```
class(photo_GovernorQuinn_likes)
```

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

8. Datatype of status\_BruceRauner

```
class(status_BruceRauner)
```

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

9. Datatype of status\_GovernorQuinn

```
class(status_GovernorQuinn)
```

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

10. Datatype of status\_BruceRauner\_likes

```
class(status_BruceRauner_likes)
```

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

11. Datatype of status\_GovernorQuinn\_likes

```
class(status_GovernorQuinn_likes)
```

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

12. Datatype of video\_BruceRauner

```
class(video_BruceRauner)
```

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

13. Datatype of video\_GovernorQuinn

```
class(video_GovernorQuinn)
```

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

14. Str of final.df

```
str(final.df)
```

```
## 'data.frame':    2 obs. of  4 variables:
##  $ Name      : Factor w/ 2 levels "Bruce Rauner",...: 2 1
##  $ Photo_likes : num  28815 73623
##  $ video_likes : num  1082 20045
##  $ Status_likes: num  13175 37377
```

15. Summary of final.df

```
summary(final.df)
```

```
##           Name      Photo_likes      video_likes      Status_likes
## Bruce Rauner:1      Min.       :28815      Min.       : 1082      Min.       :13175
## Pat Quinn   :1      1st Qu.:40017      1st Qu.: 5823      1st Qu.:19226
##           Median   :51219      Median   :10564      Median   :25276
##           Mean     :51219      Mean     :10564      Mean     :25276
##           3rd Qu.:62421      3rd Qu.:15304      3rd Qu.:31326
##           Max.     :73623      Max.     :20045      Max.     :37377
```

**Result Section**

Below is the final/cleaned data table showing number of likes on status, photos and videos of Pat Quinn and Bruce Rauner

```
final.df
```

```
##           Name Photo_likes Video_likes Status_likes
## 1      Pat Quinn      28815        1082        13175
## 2 Bruce Rauner      73623        20045        37377
```

*Table 1 shows the cleaned data collected from the data sources.*

From the above table, we can analyze that the number of likes on photos, videos and status of Bruce Rauner’s facebook page is greater than that of Pat Quinn’s (Governor Quinn). Moreover, we can make our prediction that Bruce Rauner is more likely to win the election than Pat Quinn.

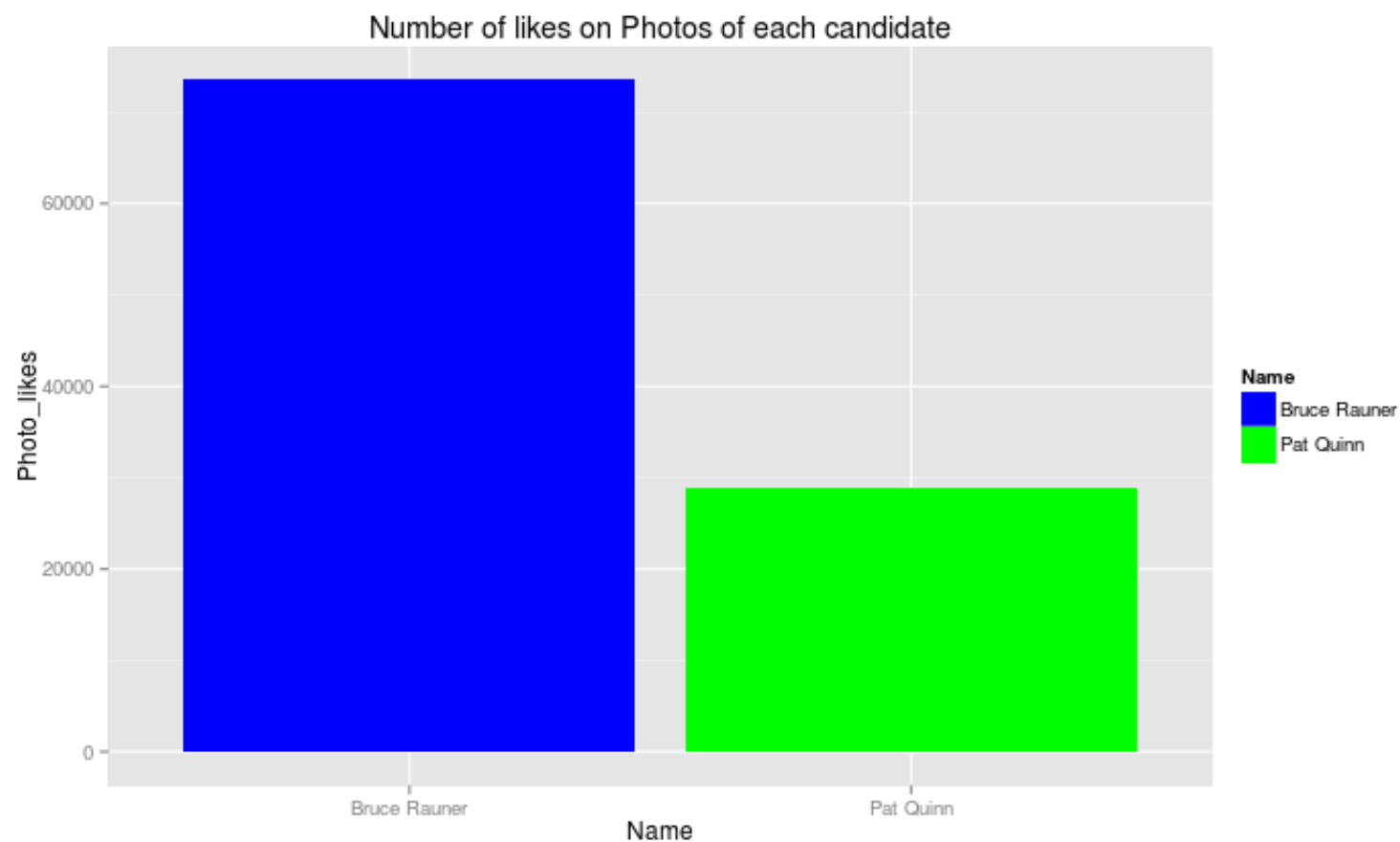
Below is the graphical analysis showing the popularity of Pat Quinn and Bruce Rauner

1. Graph to represent number of likes on Photos of each candidate’s facebook page.

ggplot2 is a graph plotting system in R. This package provides us various functionalities to create graphs through which we can showcase our analysis in more presentable form.

Using the below code, Name is presented on the X-axis and number of likes on the photos of Pat Quinn and Bruce Rauner are presented on Y-axis. geom\_bar provides us a functionality to create bar graph. Through, scale\_fill\_manual we can fill blue and green color in our bars. Moreover, using ggtitle we can provide title to our graph.

```
library(ggplot2)
ggplot(data=final.df, aes(x=Name, y=Photo_likes, fill=Name)) + geom_bar(stat="identity") + scale_fill_manual(values=c("blue",
"green")) + ggtitle("Number of likes on Photos of each candidate")
```



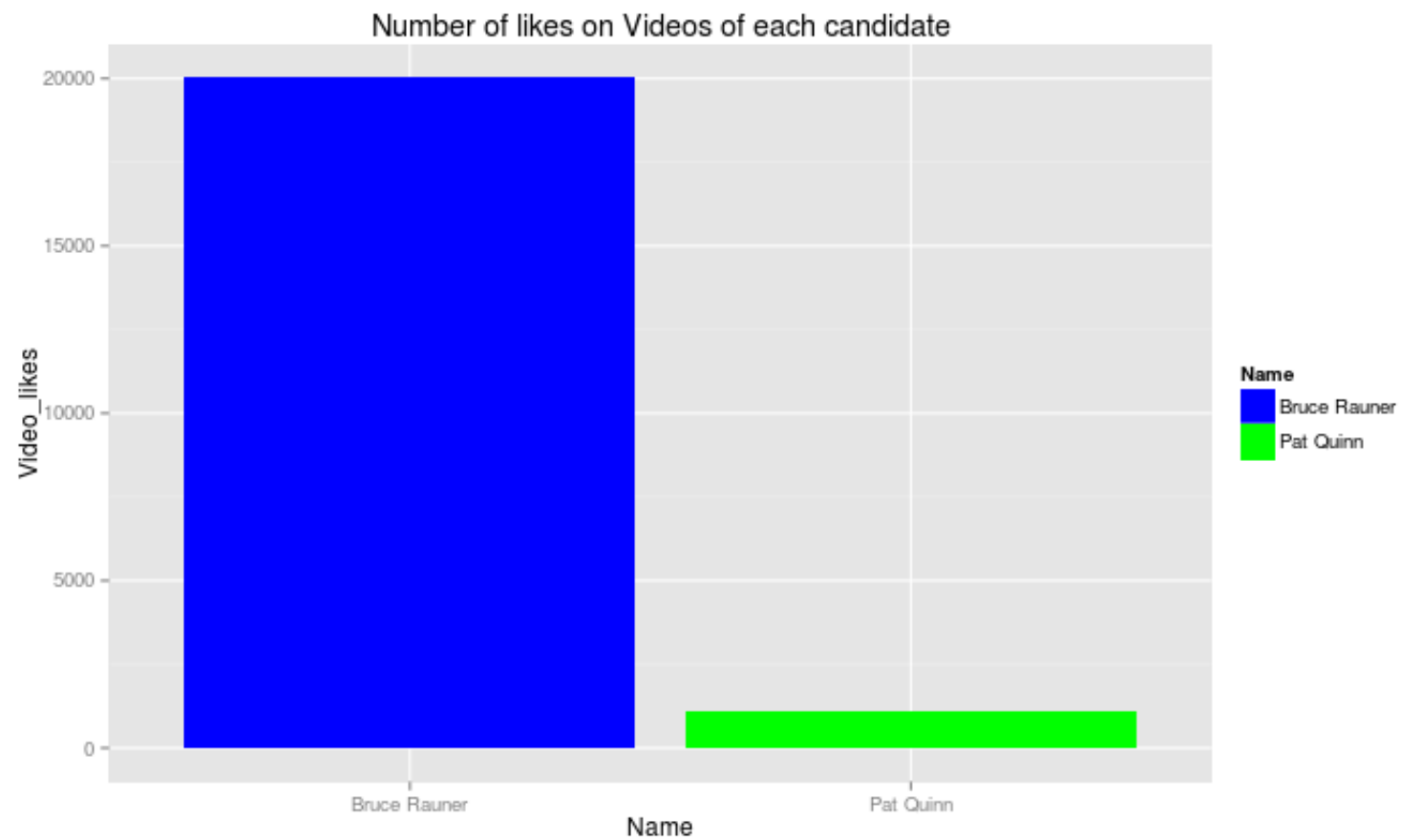
*Fig. 1 compares the number of likes on photos of each candidate.*

Above graph provides graphical representation of number of likes received on the photos of Bruce Rauner's and Pat Quinn's facebook page. Based on the analysis, we can conclude that number of likes for Bruce Rauner is nearly 2.5 times higher than that of Pat Quinn.

## 2. Graph to represent number of likes on Videos of each candidate's facebook page

Using the below code, Name is presented on the X-axis and number of likes on the video of Pat Quinn and Bruce Rauner are presented on Y-axis. `geom_bar` provides us a functionality to create bar graph. Through, `scale_fill_manual` we can fill blue and green color in our bars. Moreover, using `ggtitle` we can provide title to our graph.

```
ggplot(data=final.df, aes(x=Name, y=video_likes, fill=Name)) + geom_bar(stat="identity") + scale_fill_manual(values=c("blue", "green")) + ggtitle("Number of likes on videos of each candidate")
```



*Fig. 2 compares the number of likes on videos of each candidate.*

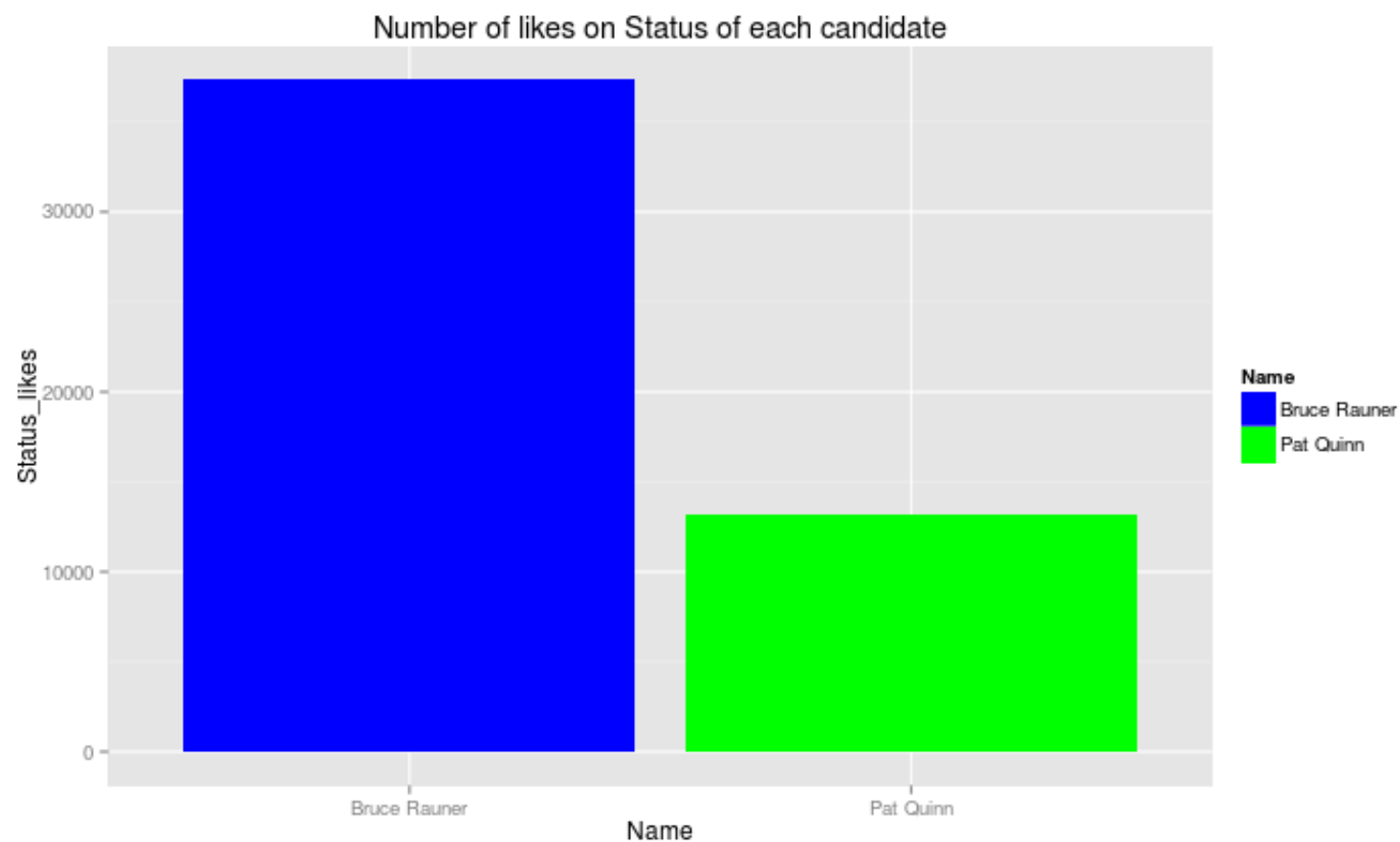
Above graph provides graphical representation of number of likes received on the videos of Bruce Rauner's and Pat Quinn's facebook page. Based on the analysis, we can conclude that number of likes for Bruce Rauner is nearly 12 times higher than that of Pat Quinn.

1. Graph to represent number of likes on Status of each candidate's facebook page

Using the below code, Name is presented on the X-axis and number of likes on the status of Pat Quinn and Bruce Rauner are presented on Y-axis. `geom_bar` provides us a functionality to create bar graph. Through, `scale_fill_manual` we can fill blue and green color in our bars. Moreover, using `ggtitle` we can provide title to our graph.

```
ggplot(data=final.df, aes(x=Name, y=Status_likes, fill=Name)) + geom_bar(stat="identity") + scale_fill_manual(values=c("blue", "green")) + ggtitle("Number of likes on Status of each candidate")
```





*Fig. 3 compares the number of likes on status of each candidate.*

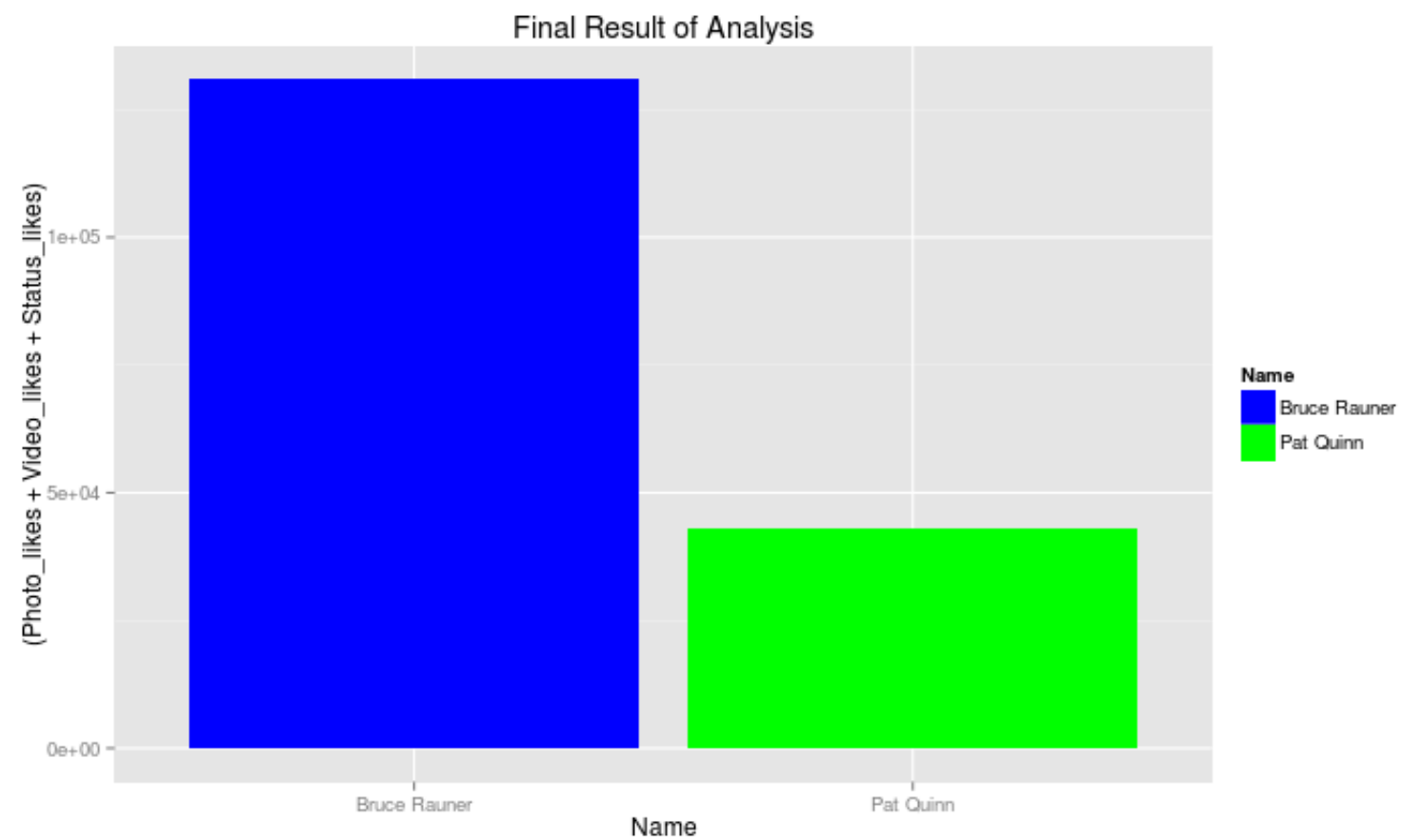
Above graph provides graphical representation of number of likes received on the status of Bruce Rauner's and Pat Quinn's facebook page. Based on the analysis, we can conclude that number of likes for Bruce Rauner is nearly 3 times higher than that of Pat Quinn.

## **Conclusion:**

Graphical Representation of final prediction:

In the below code, Name is presented on the X-axis and total number of likes on the photos, videos and status of Pat Quinn and Bruce Rauner are presented on Y-axis. `geom_bar` provides us a functionality to create bar graph. Through, `scale_fill_manual` we can fill blue and green color in our bars. Moreover, using `ggtitle` we can provide title to our graph.

```
ggplot(data=final.df, aes(x=Name, y=(Photo_likes+Video_likes+Status_likes), fill=Name)) + geom_bar(stat="identity") +
scale_fill_manual(values=c("blue", "green")) + ggtitle("Final Result of Analysis")
```



*Fig. 4 shows the final result of predictive analysis about winner of Illinois Governor's Race.*

The approach explained in this report provides a way to make use of facebook to find the popularity of each candidate standing in the election for Illinois Governor position. We can also understand the fact that how likes on photos, videos and status posted by each candidate can help us in connecting the dots between their popularity and chance of winning the election. Moreover, the tabular and graphical analysis using ggplot2 package we can understand the results more thoroughly and substantially.

Source:

1. <http://pablobarbera.com/blog/archives/3.html>
2. <http://cran.r-project.org/web/packages/Rfacebook/Rfacebook.pdf>