Act_report

May 13, 2018

1 Project Wrangle and Analyze Data: Act Report

1.0.1 *Author: Anthony T. O'Brien M.D.*

12th May 2018

2 Introduction

My analysis focused on various parts of the wrangled dataframe. Specifically, I looked at the relationship between the number of favorites relative to the number of retweets, the impact that a dog's score would have on the aformentioned relationship, a quick count of the most popular methods to post these tweets (including a quick review of the most popular dog breed), and finally a review of the scoring pattern over the time period in which the data were collected.

3 Methods

To study the relationship between the number of favorited tweets relative to the number of retweets, I used a scatterplot wiht a regression line superimposed.

Once the scatterplot was created I continued with a bubble plot which adds a third variable to give an extra dimension to the plot, this variable was the score_ratio divided by 18.

At this point a simple barplot was used to visualize the methods used to post tweets, and a count table was created to determine which were the top three dog breeds.

Finally as a continuation of the score analysis, I plotted the score ratio relative to the time the score was tweeted.

```
In [11]: from IPython.display import HTML

         HTML('''<script>
         code_show=true;
         function code_toggle() {
          if (code_show){
          $('div.input').hide();
          } else {
          $('div.input').show();
          }
          code_show = !code_show
```

```
}
$( document ).ready(code_toggle);
</script>
<form action="javascript:code_toggle()"><input type="submit" value="Click here to toggl
```

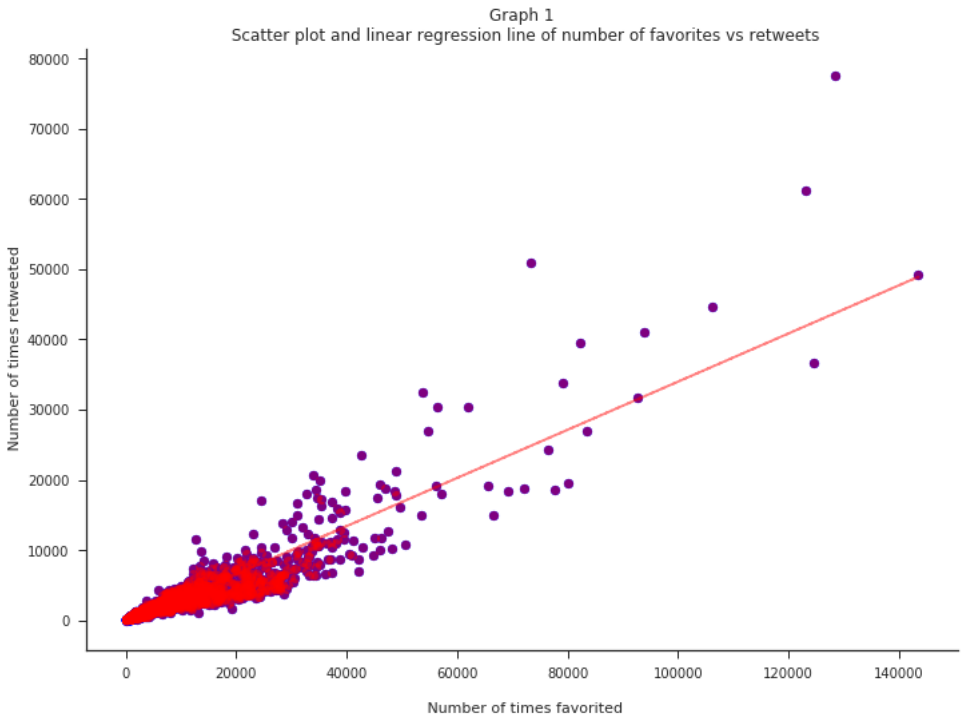
Out[11]: <IPython.core.display.HTML object>

4 Results

Graph 1 plots the relationship between the number of tweets and retweets in a scatterplot with a superimposed linear regression.

In [19]:
```
from IPython.display import Image
Image("Img/graph1.png")
```

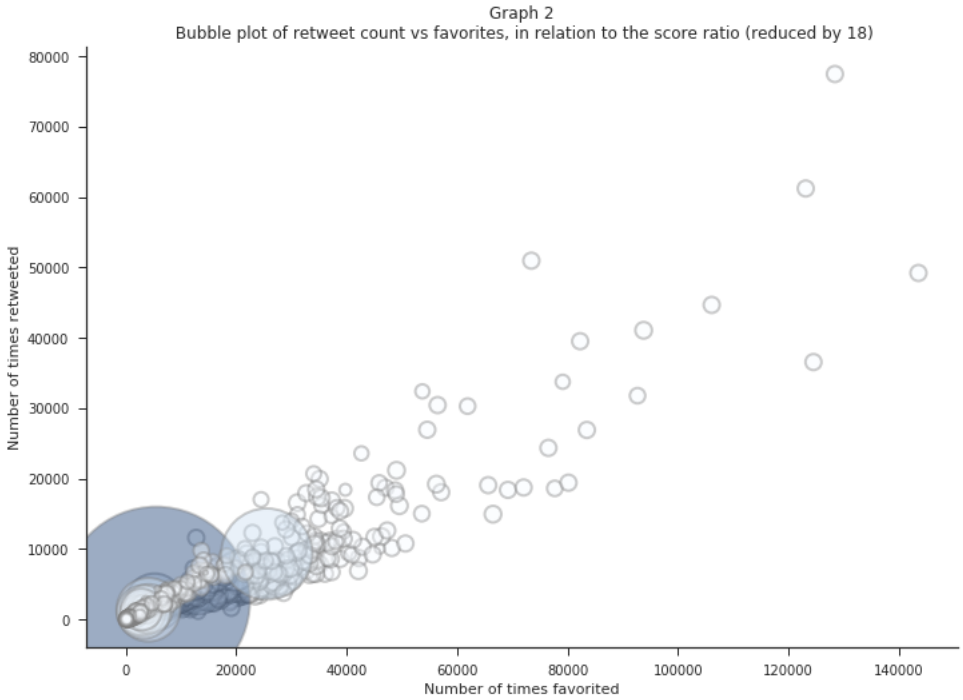
Out[19]:



Graph 2 builds on graph 1 by adding a density variable to the z axis (i.e. score ratio/18).

In [21]:
```
from IPython.display import Image
Image("Img/graph2.png")
```

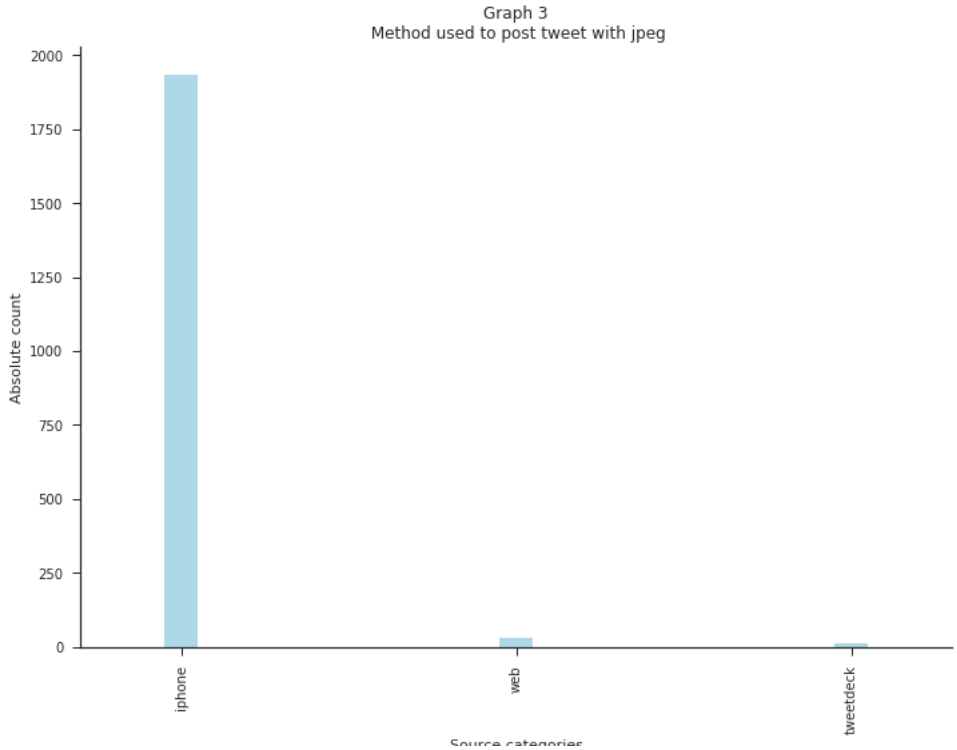
Out[21]:



Graph 3 plots a bargraph of the method for publishing tweets (i.e. iPhone, Webclient, or tweetdeck).

```
In [22]: from IPython.display import Image
         Image("Img/graph3.png")
```

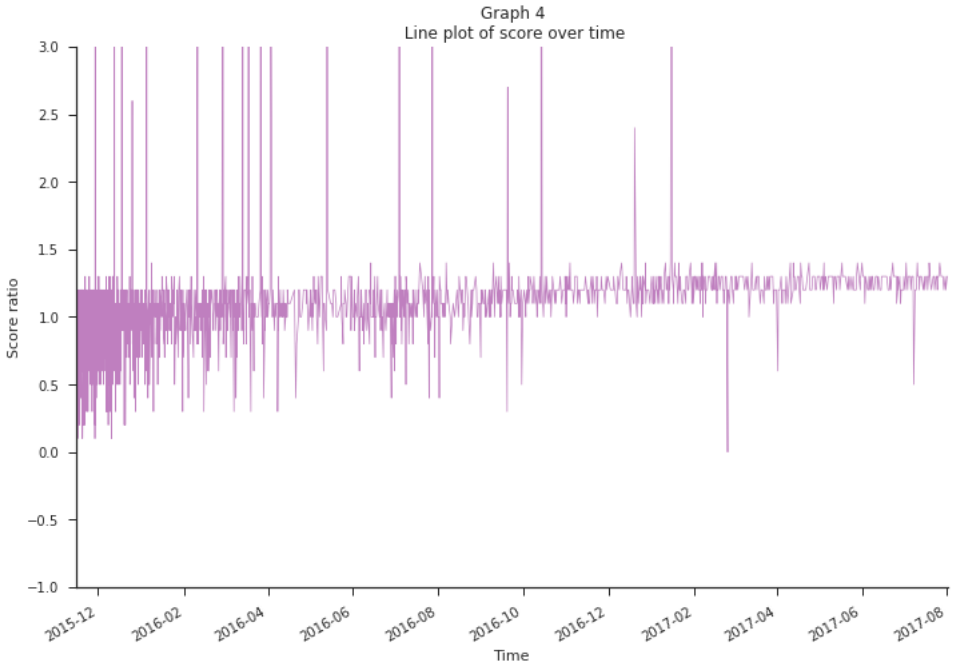
Out[22]:



Graph 4 studies the reltionship between the dog's score_ratio over the period of time in which the data was collected

```
In [16]: from IPython.display import Image
         Image("Img/graph4.png")
```

Out[16]:



5 Analysis

What we learn from these results is the following:

Relationship between favorites and retweets

- 1) There is a linear relationship between the number of favorited tweets vs the number of retweets, although this relationship becomes less obvious over 30,000 favorites.

Importance of dog score

- 2) The impact of score on the relationship between favorited and retweeted tweet is negligable.
- 3) As time progresses the score succumbs to regression to the mean and becomes less relevant as compared to the beginning. That is to say that the scores tend to 1

Method of tweeting

- 4) The iphone was the most preferred method for perfoming tweets in this data set, followed by WebClient and lastly tweetdeck

Most posted dogs

- 5) The top three dog breeds tweeted in this data set were the Golden Retriever, the Labrador and the Pembroke