act_report

June 21, 2018

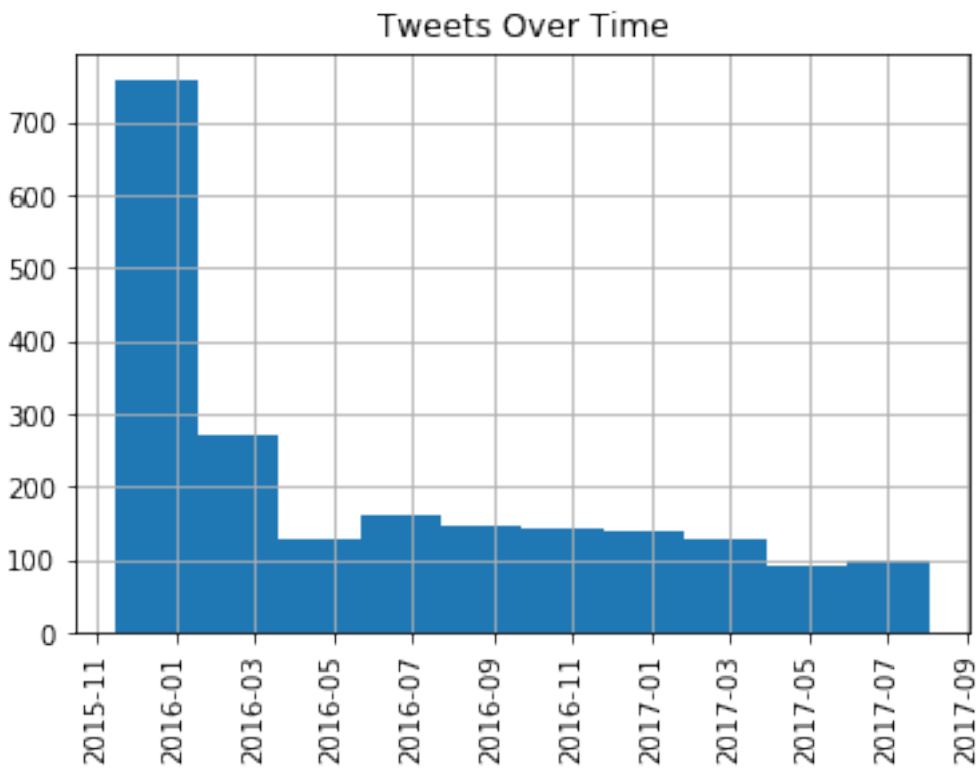
```
In [6]: import json
        import matplotlib.pyplot as plt
        import numpy as np
        import pandas as pd
        import requests
        import seaborn as sns
        import tweepy

        %matplotlib inline
```

```
In [7]: df = pd.read_csv('twitter_archive_master.csv')
```

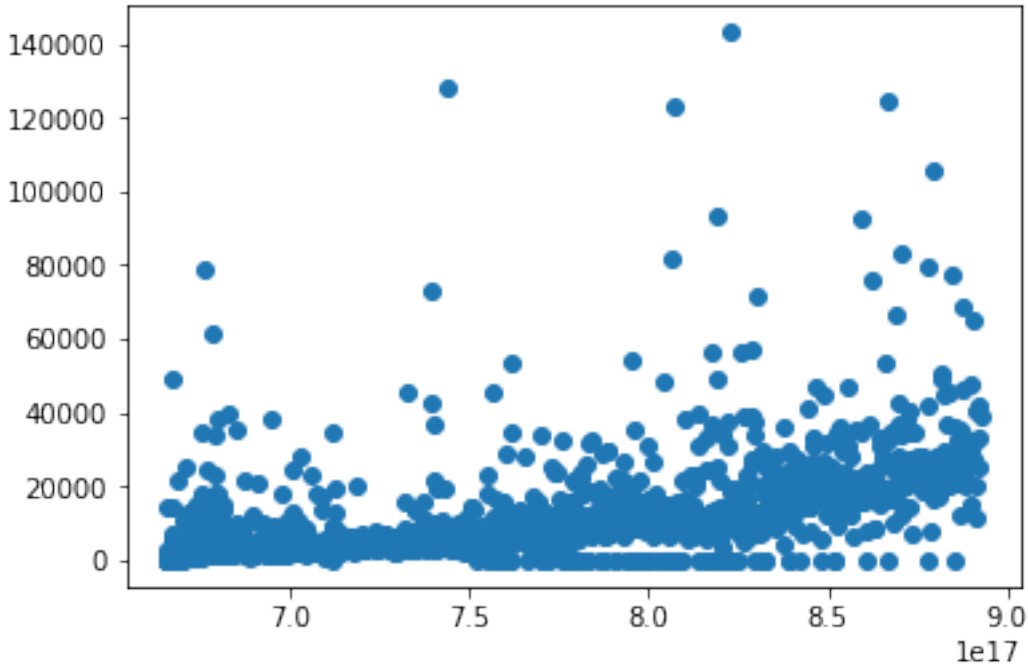
First, we'll take a look at the activity of WeRateDogs. We can see that the account started off with a very high number of tweets, but soon found a sustainable rhythm. For the past few years, the number of tweets has been slowly trending downward.

```
In [18]: df['timestamp'].astype('datetime64[ns]').hist(xrot=90);
         plt.title('Tweets Over Time');
```



This is an interesting finding that begs for some follow up. What caused the enormous drop in activity, and why did the number of tweets stabilize? When we dive deeper, we can see that despite the number of tweets going down, the average popularity of each tweet goes up. We would want to explore this from more angles before making absolute statements, but we can tentatively conclude that the number of tweets went down because the account grew more popular and simply didn't need to tweet as much.

```
In [19]: plt.scatter(x=df['tweet_id'], y=df['favorite_count']);
```



WeRateDogs is known for its humerous rating of dogs, specifically that the ratings are almost always above 100%. We can combine the numerators and denominators to see what the average rating looks like. We can see that, despite hyperbolous ratings, the average tends to be only slightly higher than 1. The max is a whopping 177.

```
In [13]: df['score'] = df['rating_numerator'] / df['rating_denominator']

         df['score'].describe()

Out[13]: count    2069.000000
         mean        1.169591
         std         3.992727
         min         0.000000
         25%         1.000000
         50%         1.100000
         75%         1.200000
         max       177.600000
         Name: score, dtype: float64
```

I would also like to look at the descriptions to see if any are favored. We can see that most dogs are classified as 'puppers' (I would have guessed 'doggos'. 'Floofer' is hardly ever used at all.

```
In [14]: df['dog_description'].value_counts()

Out[14]: pupper     231
         doggo       75
```

```
puppo       29
floofer      3
Name: dog_description, dtype: int64
```

In a similar vein, and for my own personal curiosity, I would like to see how popular each dog name is! We can see that names like Charlie, Penny, and Oliver are common, as we could probably expect. The names with only one instance can be pretty fun, like Michelangelope, Chuq, and Enchilada.

```
In [17]: df['name'].value_counts()
```

```
Out[17]: Charlie           11
         Penny             10
         Oliver            10
         Lucy              10
         Tucker            10
         Cooper            10
         Sadie              8
         Winston            8
         Bo                 8
         Lola               8
         Daisy              7
         Toby               7
         Stanley            6
         Koda               6
         Milo               6
         Bella              6
         Dave               6
         Rusty              6
         Jax                6
         Scout              6
         Bailey             6
         Buddy              5
         Chester            5
         Oscar              5
         Alfie              5
         Larry              5
         Louis              5
         Leo                5
         Derek              4
         Gus                4
                           ..
         Charl              1
         Ralphé             1
         Snoopy             1
         Robin              1
         Amy                1
         Dante              1
```

```
Tassy              1
Brat               1
Willem             1
Clifford           1
Pinot              1
Michelangelope     1
Kaiya              1
Blue               1
Edd                1
Bodie              1
Aja                1
Chuq               1
Jo                 1
Jaspers            1
Colin              1
Shnuggles          1
Steve              1
Enchilada          1
Anakin             1
Tessa              1
Mya                1
Charleson          1
Sundance           1
Aiden              1
Name: name, Length: 913, dtype: int64
```