Analyzing and Visualizing WeRateDogs

In this project, I will be analyzing data associated with the witter_archive_master.csv. The dataset contains 1994 rows and 21 columns. Relevant data to be used in the following analysis include the following variables:
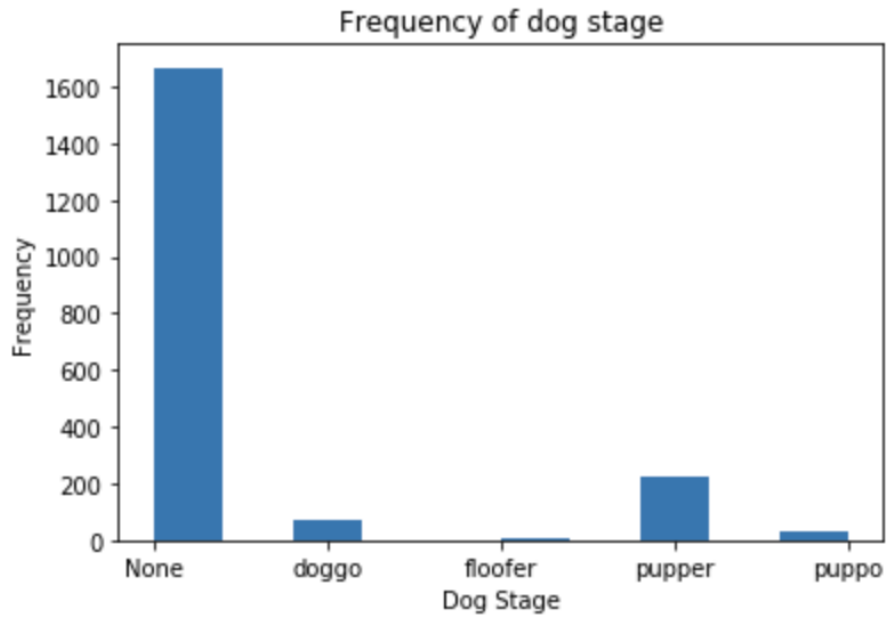
- dog stage (doggo, floofer, pupper, puppo)
- rating (rating_numerator / rating_denominator)
- name
- retweet count
- favorite count
- timestamp

In particular, I will be interested in the following questions:

- What's the common dog stage?
- Which dog stage has highest rating?
- What's the top 10 dog's name?
- What's the relationship between tweet count and favorite count?
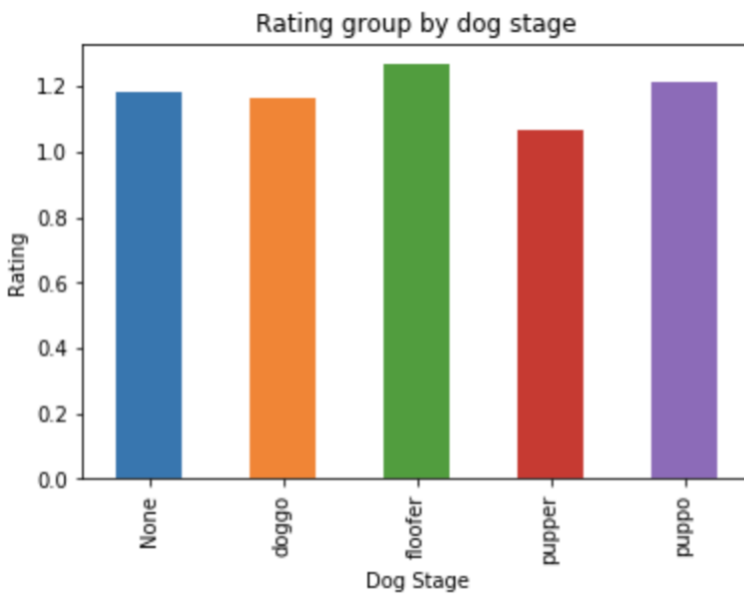- Which month has most retweet and like?

Question 1: What's the common dog stage?

```
pupper    223
doggo      72
puppo      28
floofer     3
```

Frequency of dog stage

From the output, we could see the most common dog stage is popper which has 1668 dogs, after that is doggo which has 72 dogs, and poppo is 28, the least dog stage is floofer which has only 3 dogs. Alternativity, most dog in this dataset has no information about dog stage, which is the highest bar in the histogram.
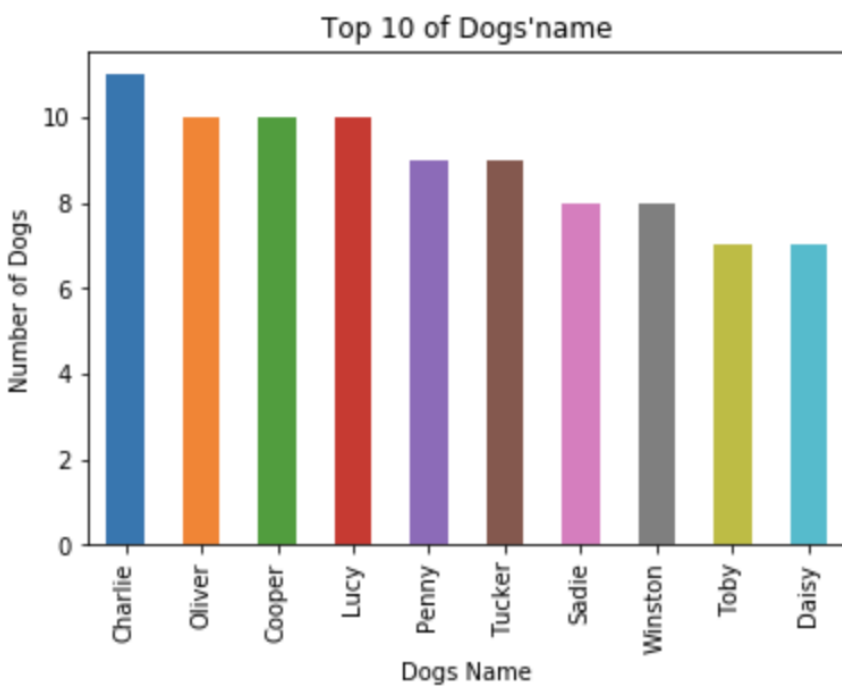
Question 2: Which dog stage has highest rating?

Rating group by dog stage

From the graph, it is hard to see the difference of rating between four dog stage, the rating for floofer is slightly higher than others.
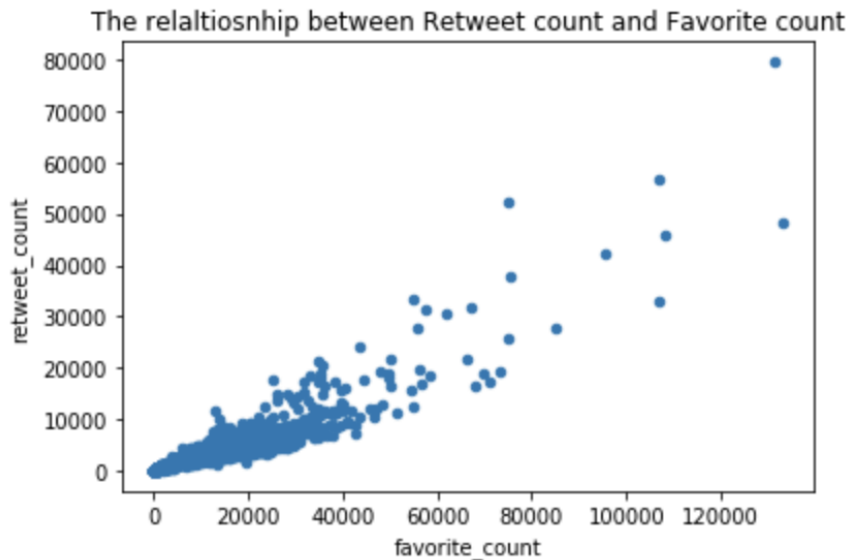
Question 3: What's the top 10 dog's name?

Charlie   11
Oliver   10
Cooper   10
Lucy   10
Penny   9
Tucker   9
Sadie   8
Winston   8
Toby   7
Daisy   7

Top 10 of Dogs'name

Expect dogs do not have a name, the most common dog's name is Charlie, there are 11 dogs have the name, Charlie. There are 10 dogs called Lucy, Oliver and Cooper. The name of Tucker and Penny both have 9 counts. There are 8 dogs called Sadie and Winston. The last place of top 10 names is Daisy and Lola, which has 7 dogs called that name.

Question 4: What's the relationship between tweet count and favorite count?

From the output, there is a positive correlation between how much a post was retweeted and favorite counts, the more favorite count, the more retweet count. People through retweet the post to show their like for the retweet post.

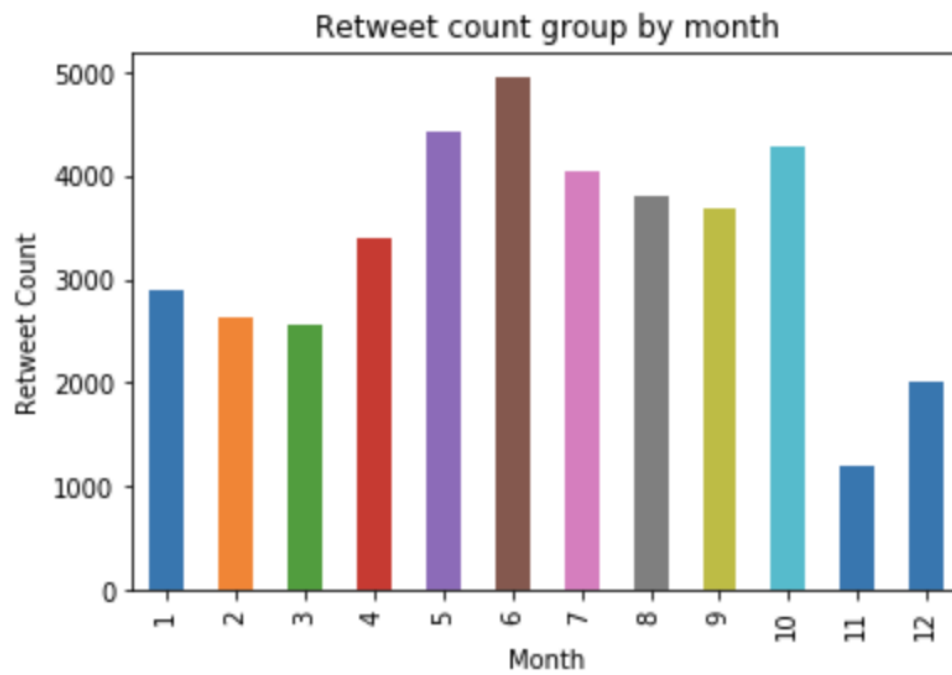**Question 5: Which month has most retweet and like?**
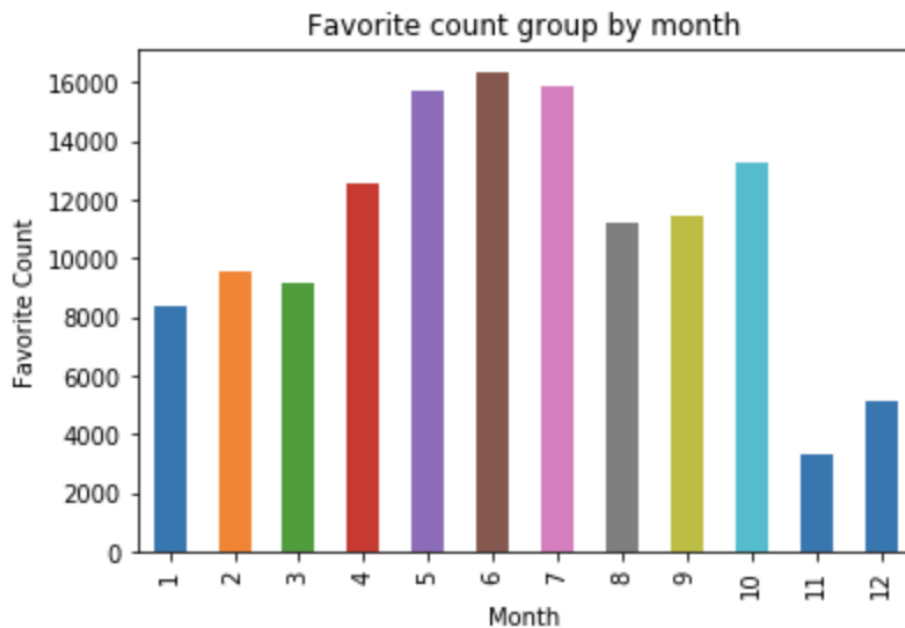
Retweet count group by month:

```
1    2890.612766
2    2631.765714
3    2554.708333
4    3399.494737
5    4418.320000
6    4951.056000
7    4051.255474
8    3803.393443
9    3687.285714
10   4274.276923
11   1196.636103
12   2017.610451
```

Favorite count group by month:

```
1    8389.702128
```

2	9556.234286
3	9141.404762
4	12534.168421
5	15722.570000
6	16341.232000
7	15858.138686
8	11170.901639
9	11435.476190
10	13294.169231
11	3289.808023
12	5145.021378

Retweet count group by month

Favorite count group by month

From the output, people most frequently retweet and give a like for a post in May and June. The post retweet in May has 4418.32 times, and 4951.056 times in June. People give a like for a post in May is Favourite count is 15722.57 times and 16341.232 times in June. Compare retweet and favourite count, people would more likely add a post as favourite, rather than retweet it.

This report evaluates the most common dog stage is popper, after that is doggo and poppo, the least dog stage is floofer. There is no big difference in the rating between the dog stage, the least dog stage has a slightly higher rating than the other stage. Most people called their dogs *Lucy, Oliver and Cooper.* There is a positive correlation between how much a post was retweeted and favourite counts, the more favourite count, the more retweet count. People would more likely add a post as the favourite, rather than retweet it and in May and June, people would more frequently retweet and give a like for a post.