Data Visualization

The dataset that I visualized is the tweet archive of Twitter.user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's.
dogs with a humorous comment about the dog. These ratings always have a denominator of 10.The numerators, though? Always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

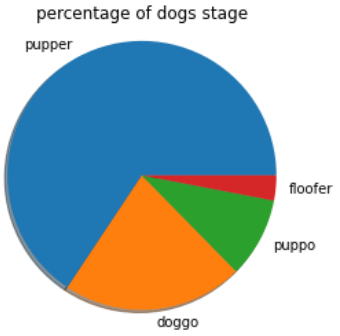
(1) percentage of dogs stage with 2 different ways in the visualizing:

1-using:pie : to visualize the 4 dogs stage

(1) percentage of dogs stage with 2 diffrenent ways in visualizing:

In [89]:
```
plt.pie(dogs_stage_visualization,
        labels=['pupper','doggo','puppo','floofer'],
        shadow=True)
plt.title('percentage of dogs stage')
plt.axis('equal')
```

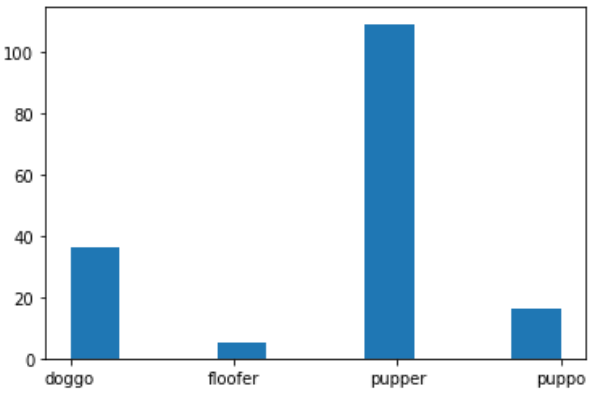
Out[89]:
```
(-1.1074088151441537,
 1.1003528049759264,
 -1.1182393450067951,
 1.1048587908238106)
```



2-using:hist : to visualize the 4 dogs stage

```
In [90]: plt.hist(twitter_archive_clean.dogs_stage)
```

```
Out[90]: (array([ 36.,   0.,   0.,   5.,   0.,   0., 109.,   0.,   0.,  16.]),
          array([0. , 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.1, 2.4, 2.7, 3. ]),
          <a list of 10 Patch objects>)
```



So the most dog stage is pupper.

(2)Relation between retweet_count and favorite_count Using:scatter

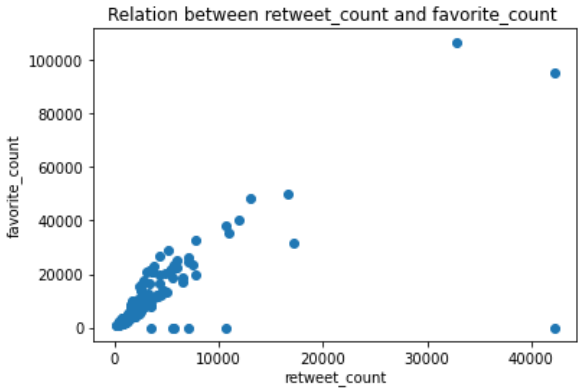
X axis →retweet_count

Y axis →favorite_count

(2)Relation between retweet_count and favorite_count

```
In [91]: plt.scatter(twitter_archive_clean.retweet_count,twitter_archive_clean.favorite_count)
         plt.xlabel('retweet_count')
         plt.ylabel('favorite_count')
         plt.title('Relation between retweet_count and favorite_count ')
```

```
Out[91]: Text(0.5, 1.0, 'Relation between retweet_count and favorite_count ')
```



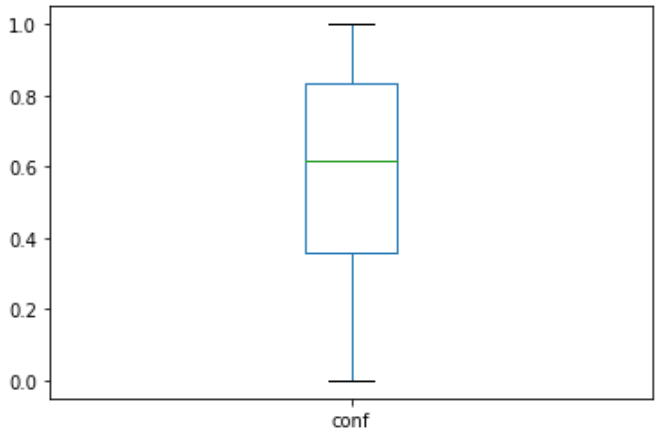
So the relation between them is a positive relationship (increase together)

(3)show the analysis of conf Using box plot:

(3)show the analysis of conf:

```
In [92]: twitter_archive_clean.conf.plot(kind='box')
```

```
Out[92]: <matplotlib.axes._subplots.AxesSubplot at 0x24cb210dfa0>
```



The max value is almost close to 1.0
The mean is 0.6
The min value is almost close to 0
The 25% value closes to 0.4
The 75% value closes to 0.8

(4) percentage of best dogs type (12 type):

Using: pie to visualize the 12 type

(4) percentage of best dogs type (12 type):

In [95]:
```
plt.pie(most_dogs_type,
       labels=['Golden_Retriever','Pug','Pomeranian','Labrador_Retriever','Pembroke','German_Shepherd',
            'Chihuahua', 'Toy_Poodle','Border_Collie', 'French_Bulldog','Staffordshire_Bullterrier','Eskimo_Dog'],
       shadow=True)
plt.title('percentage of best dogs types')
plt.axis('equal')
```

Out[95]:
```
(-1.1000000094094384,
 1.1000000175915503,
 -1.1000000130913898,
 1.100000049092735)
```