Analysis & Visualization

Based on the master data frame which consists of the three data frames, the following are the analysis:

1. The ratings of the dogs are mostly between 0.8 to 1.4, with many ratings around 1.15-1.25. The graph is righly skewed, with more ratings are inclined towards the higher ratings. This makes sense as more tweets would be done for high rated dogs. The graph is bimodal at ratings around 1.0 and 1.2.
2. The newly engineered attribute 'status' may contain values from doggo, floofer, puppo, pupper and multiple. To avoid any confusion, the values were extracted from the text column (in the original dataframe). The most common tweeted dog is generally a pupper which constitute about 69%. While doggo and puppo were 16% and 12% respectively. Least tweeted dogs were the floofers. Also the multiple category has around 3% of the dogs which were mostly pupper and doggo. Please refer to the pie chart below.
3. As expected the scatter plot between retweeted count and favorite count is a straight line. There is a high correlation factor between them. It is calculated to be .921 where 1.0 is maximum. We can't draw a casuality without the proper use of statistics or machine learning techniques, but it can be seen that tweets that retweet and favorite count are proportional.
4. Based on the category of dog status, the mean rating of a puppo is the maximum. It is followed by duggo, floofer and pupper. Similar numbers are seen in terms of mean of favorites and retweet counts.
5. Finally, based on the dog breed the average ratings are high for Saluki, Tibetan Mastiff and Briard. Saluzi and Tibetian Mastiff also have high average retweet and favorite counts, and in general very popular on Twitter. On the other hand, Japanese Spaniel and black and tan coonhound are among the least tweeted (rated/retweeted) dogs.

In [10]:
```
import pandas as pd
twitter_archive_master = pd.read_csv('twitter_archive_master.csv')
twitter_archive_master.status_of_dog.value_counts().plot(kind='pie', figsize =
(5,5), autopct='%1.0f%%');
```