

Insights

- 1.主人对狗狗评分的平均数为12，但由于有较大的异常值（如1776），所以参考平均数不够准确。
- 2.评分的中位数为11，上下四分位数分别为10和12，说明主人给狗狗的评分大部分都超过了十分制的满分，主人对自己的狗狗普遍评价很高。
- 3.参考p1_conf，对狗的图片的识别置信度平均数为0.59，最高可达到1。对于图片的识别不是特别准确，存在误判。同时一些图片中的狗较难分辨。

点赞数最高的狗的照片是：



转发最高



visualization

由于主人给狗狗的评分存在过高的分数，如1776，666等，所以将超过10分过多的评分视为异常值。根据评分的上下四分位数（10和12），根据Tukey's test方法，令 $K=3$ ，计算出接受极端异常值的区间上限 $12+K*(12-10)=18$ 。所以选择18作为上限，作出评分的直方图。可以看到主人对狗狗的评分主要集中在8到14之间，评分低于8分的情况比较少。

In [100]:

```
twitter_archive_master[twitter_archive_master.rating_numerator<18]['rating_numerator'].plot(kind='hist')
```

Out[100]:

<matplotlib.axes._subplots.AxesSubplot at 0x110c30a58>

