

Notebook

March 4, 2019

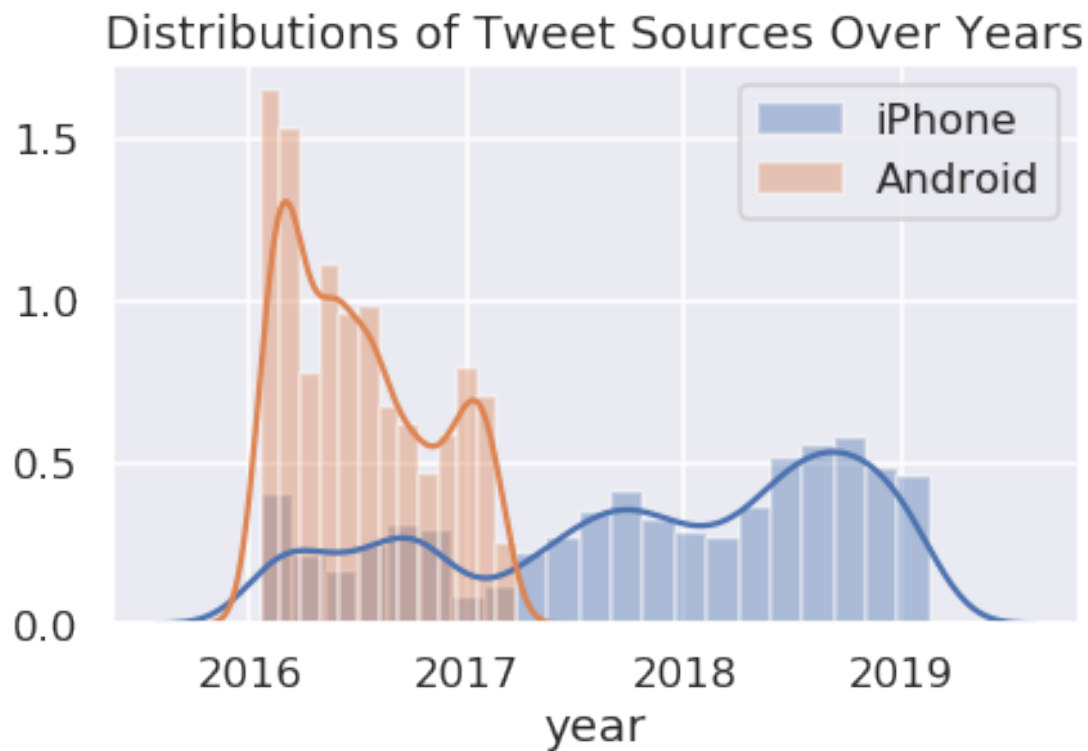
Now, use `sns.distplot` to overlay the distributions of Trump's 2 most frequently used web technologies over the years. Your final plot should look like:

```
In [60]: popular_dev = trump['source'].value_counts().head(2).index

for item in popular_dev:
    filterItem = trump['source'] == item
    sns.distplot(trump[filterItem]['year'], label=item[12:])

plt.title('Distributions of Tweet Sources Over Years')
plt.legend()
```

Out[60]: <matplotlib.legend.Legend at 0x7f56045ec438>



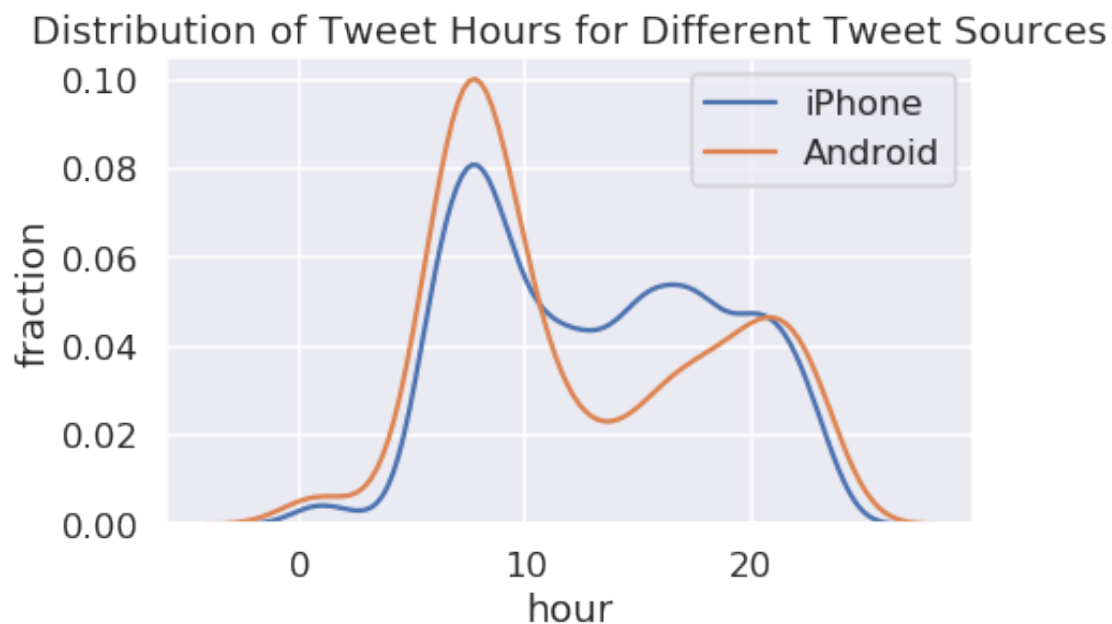
0.0.1 Question 4b

Use this data along with the seaborn distplot function to examine the distribution over hours of the day in eastern time that trump tweets on each device for the 2 most commonly used devices. Your plot should look similar to the following:

```
In [62]: ### make your plot here
popular_dev = trump['source'].value_counts().head(2).index
for device in popular_dev:
    filterItem = trump['source'] == device
    sns.distplot(trump[filterItem]['hour'], label = device[12:], hist=False)

plt.title('Distribution of Tweet Hours for Different Tweet Sources')
plt.xlabel('hour')
plt.ylabel('fraction')
plt.legend()
```

Out[62]: <matplotlib.legend.Legend at 0x7f5604e060b8>



0.0.2 Question 4c

According to [this Verge article](#), Donald Trump switched from an Android to an iPhone sometime in March 2017.

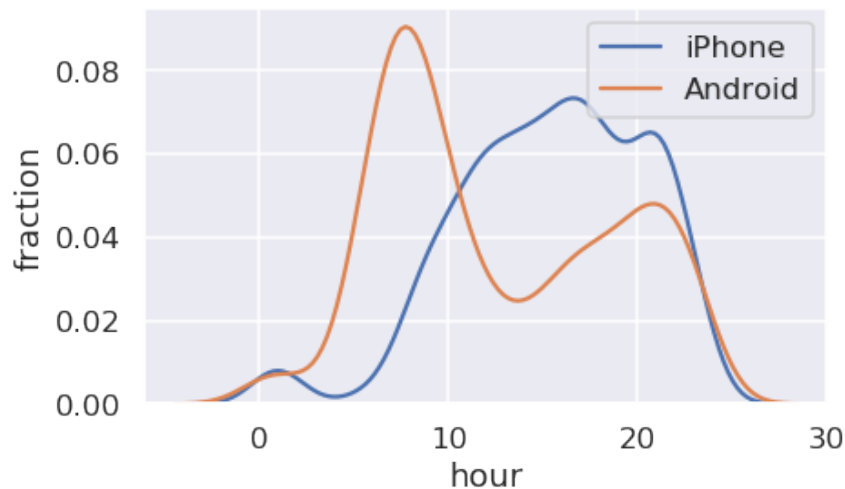
Let's see if this information significantly changes our plot. Create a figure similar to your figure from question 4b, but this time, only use tweets that were tweeted before 2017. Your plot should look similar to the following:

```
In [76]: ### make your plot here
        for device in popular_dev:
            filterItem = (trump['source'] == device) & (trump['year'] < 2017)
            sns.distplot(trump[filterItem]['hour'], label = device[12:], hist=False)

        plt.title('Distribution of Tweet Hours for Different Tweet Sources (pre-2017)')
        plt.xlabel('hour')
        plt.ylabel('fraction')
        plt.legend()
```

Out[76]: <matplotlib.legend.Legend at 0x7f5604ebdf60>

Distribution of Tweet Hours for Different Tweet Sources (pre-2017)



0.0.3 Question 4d

During the campaign, it was theorized that Donald Trump's tweets from Android devices were written by him personally, and the tweets from iPhones were from his staff. Does your figure give support to this theory? What kinds of additional analysis could help support or reject this claim?

The figure does support the theory that during the campaign, Trump used an Android phone to tweet messages while most of his staff used iPhones, based on the increased activity at around hour 8.

0.1 Question 6

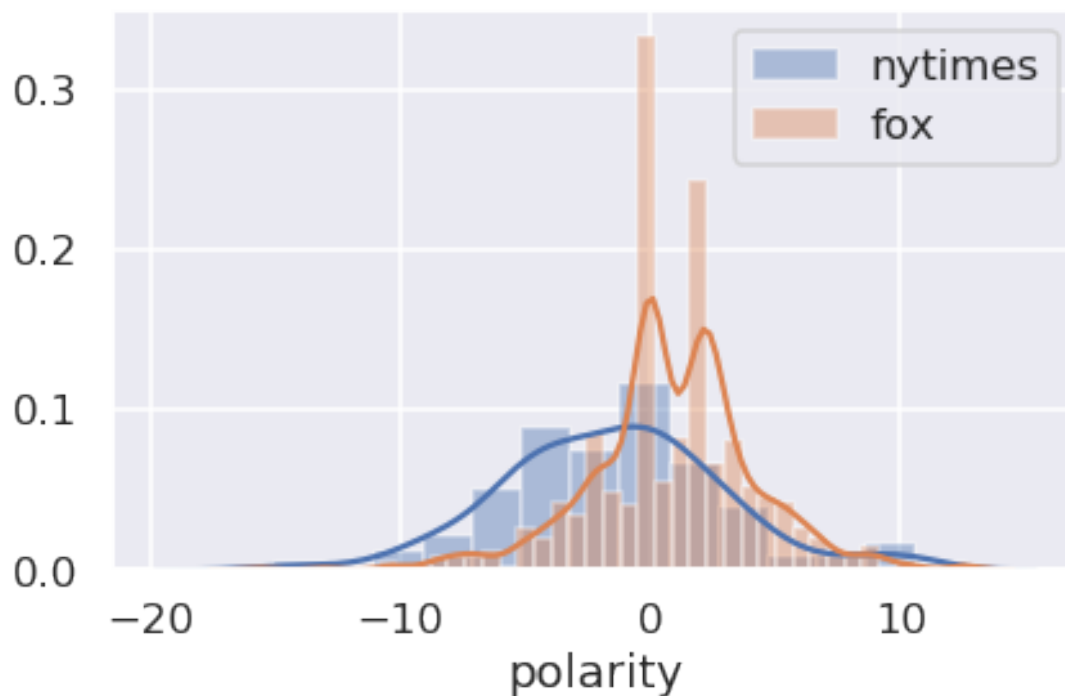
Now, let's try looking at the distributions of sentiments for tweets containing certain keywords.

0.1.1 Question 6a

In the cell below, create a single plot showing both the distribution of tweet sentiments for tweets containing nytimes, as well as the distribution of tweet sentiments for tweets containing fox.

```
In [104]: sns.distplot(trump[trump['text'].str.lower().str.contains("nyt")]['polarity'], label = 'nytim  
sns.distplot(trump[trump['text'].str.lower().str.contains("fox")]['polarity'], label = 'fox')  
plt.legend()
```

```
Out[104]: <matplotlib.legend.Legend at 0x7f5604029240>
```



0.1.2 Question 6b

Comment on what you observe in the plot above. Can you find other pairs of keywords that lead to interesting plots? (If you modify your code in 6a, remember to change the words back to `nytimes` and `fox` before submitting for grading).

Clearly, the president appears to say more positive things about Fox than the New York Times. It also appears that the president appears to say more positive things about Fox than CNN.