

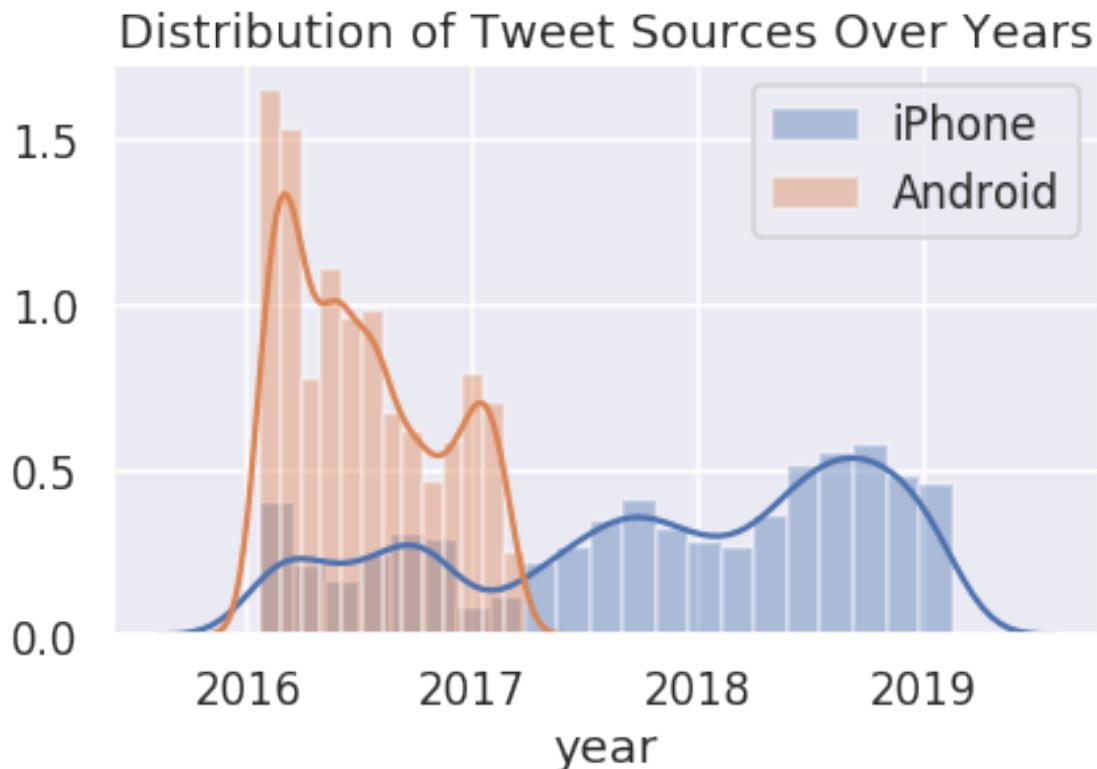
Notebook

July 11, 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 [16]: sns.distplot(trump[(trump['source'] == 'Twitter for iPhone')]['year'],label = 'iPhone')
sns.distplot(trump[(trump['source'] == 'Twitter for Android')]['year'],label = 'Android')
plt.legend()
plt.xlabel('year')
plt.title('Distribution of Tweet Sources Over Years')
plt.yticks(np.arange(0, 1.75, 0.5))

Out[16]: ([<matplotlib.axis.YTick at 0x7f376131cda0>,
<matplotlib.axis.YTick at 0x7f376131c588>,
<matplotlib.axis.YTick at 0x7f3761321390>,
<matplotlib.axis.YTick at 0x7f37612c6208>],
<a list of 4 Text yticklabel objects>)
```



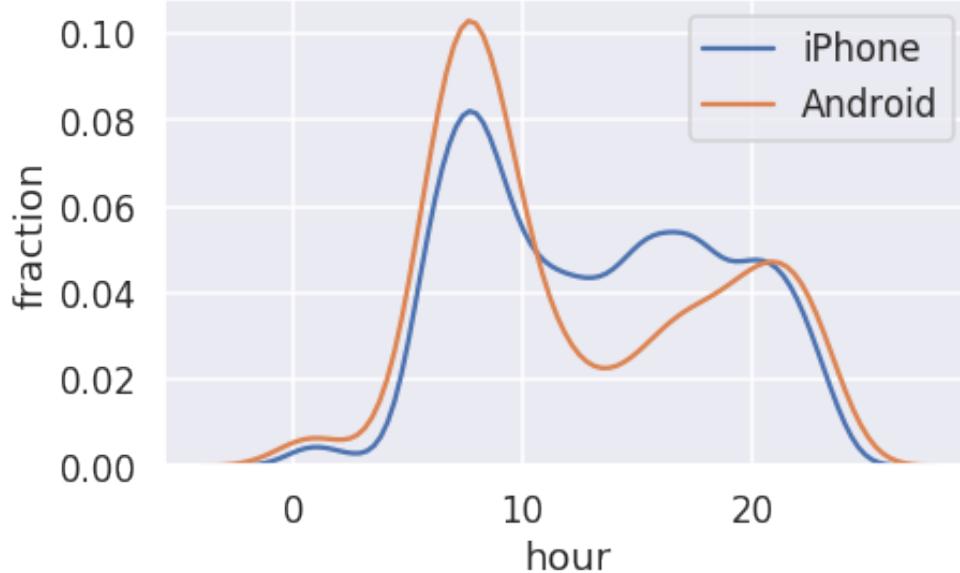
0.0.1 Question 4b

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 [21]: ### make your plot here
sns.distplot(trump[(trump['source'] == 'Twitter for iPhone')]['hour'],label = 'iPhone',hist=False)
sns.distplot(trump[(trump['source'] == 'Twitter for Android')]['hour'],label = 'Android',hist=False)
plt.legend()
plt.xlabel('hour')
plt.ylabel('fraction')
plt.title('Distribution of Tweet Hours for Different Tweet Sources')
plt.yticks(np.arange(0, 0.12, 0.02))

Out[21]: ([<matplotlib.axis.YTick at 0x7f37612609e8>,
<matplotlib.axis.YTick at 0x7f3761260278>,
<matplotlib.axis.YTick at 0x7f3761254ac8>,
... Omitting 0 lines ...
<matplotlib.axis.YTick at 0x7f376121d8d0>,
<matplotlib.axis.YTick at 0x7f376121ddd8>],
<a list of 6 Text yticklabel objects>)
```

Distribution of Tweet Hours for Different Tweet Sources



0.0.2 Question 4c

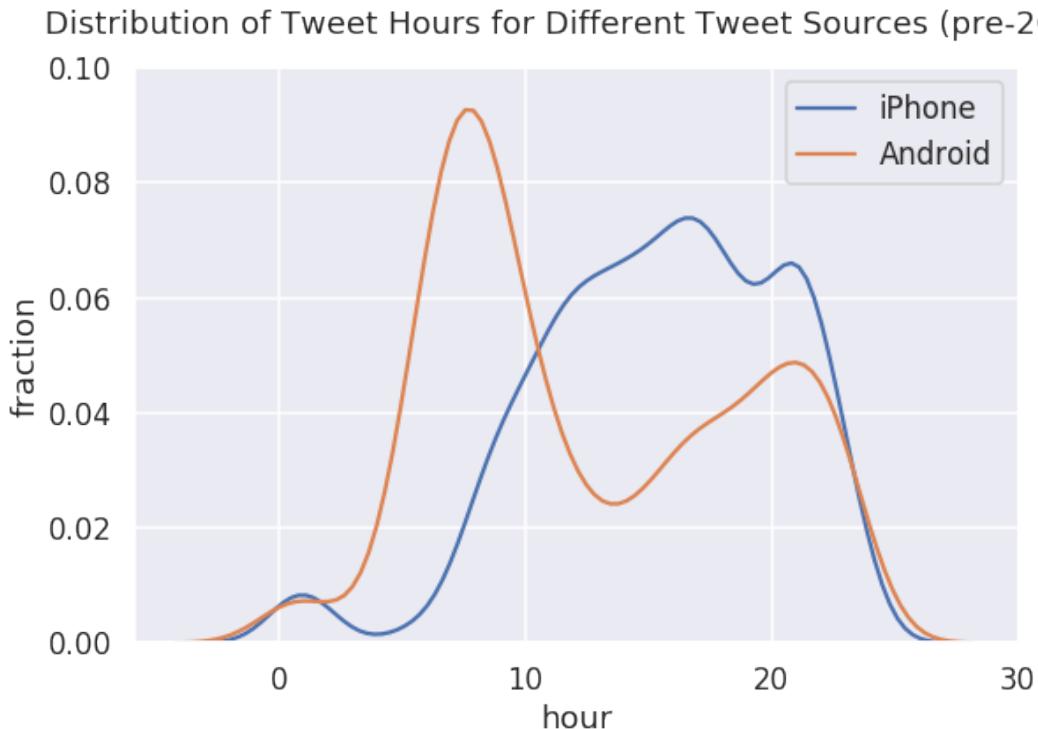
According to [this Verge article](#), Donald Trump switched from an Android to an iPhone sometime in March 2017.

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 [22]: ### make your plot here
plt.figure(figsize = (8,6))
trump_pre = trump[trump['year']<2017]

sns.distplot(trump_pre[(trump_pre['source'] == 'Twitter for iPhone')['hour'],label = 'iPhone']
sns.distplot(trump_pre[(trump_pre['source'] == 'Twitter for Android')['hour'],label = 'Android'
plt.legend()
plt.xlabel('hour')
plt.ylabel('fraction')
plt.title('Distribution of Tweet Hours for Different Tweet Sources (pre-2017)',y = 1.04)
plt.xticks(np.arange(0, 32, 10))
plt.yticks(np.arange(0, 0.12, 0.02))

Out[22]: ([<matplotlib.axis.YTick at 0x7f37611ccf28>,
<matplotlib.axis.YTick at 0x7f37611cc7b8>,
<matplotlib.axis.YTick at 0x7f37611bb470>,
... Omitting 0 lines ...
<matplotlib.axis.YTick at 0x7f3761132e10>,
<matplotlib.axis.YTick at 0x7f376113c320>],
<a list of 6 Text yticklabel objects>)
```



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?

From the figure, Donald Trump's tweets from Android devices were usually written in the morning between 5 and 10, while his tweets from iPhones were written in the afternoon often. This figure doesn't give support to this theory, it just shows that Android devices were often used to send tweets in the morning and iPhone devices were often used to send tweets in the afternoon.

Maybe from the text we could find the different idiomatic expressions with different devices and guess that could because the tweets are written by Trump and his staff. Or we should look for more extra information from other sources that support or reject this claim.

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 [74]: plt.figure(figsize = (8,5))
nytimes = trump[trump['no_punc'].str.contains('nytimes')]['polarity']
sns.distplot(nytimes,label = 'nytimes')
fox = trump[trump['no_punc'].str.contains('fox')]['polarity']
sns.distplot(fox,label = 'fox')
plt.legend()
plt.ylabel('frequency')
plt.title('Distribution of Sentiments for Tweets Containing nytimes and fox')
```

```
Out[74]: Text(0.5, 1.0, 'Distribution of Sentiments for Tweets Containing nytimes and fox')
```

