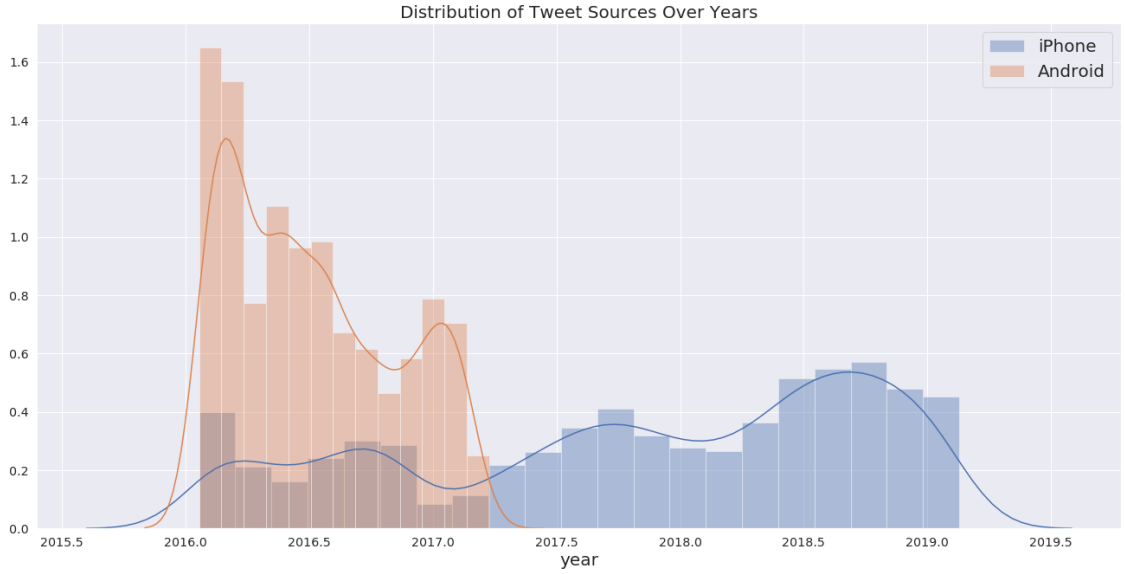
Notebook

July 12, 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:

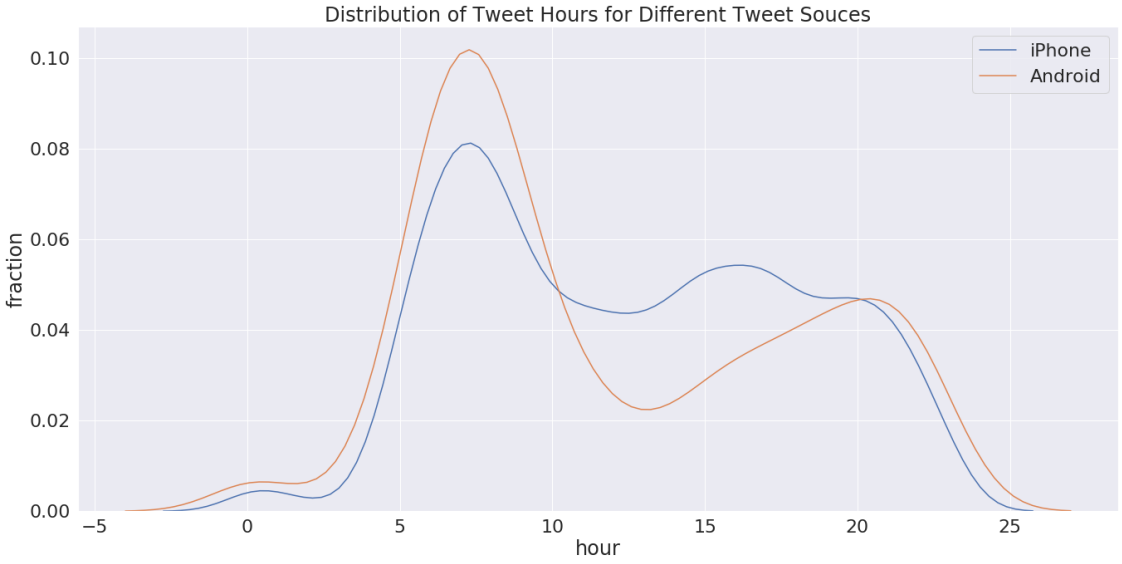
```
In [453]: iphone=trump[trump['source']=='Twitter for iPhone']['year']
          android=trump[trump['source']=='Twitter for Android']['year']
          sns.distplot(iphone,label='iPhone')
          sns.distplot(android,label='Android')
          plt.title('Distribution of Tweet Sources Over Years',fontsize=20)
          plt.xlabel('year',fontsize=20)
          plt.legend(prop={'size':20})
          sns.set(font_scale=2)
```



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 [458]: ### make your plot here
          iphone_hour = trump[trump['source']=='Twitter for iPhone']['est_time'].dt.hour
          android_hour = trump[trump['source']=='Twitter for Android']['est_time'].dt.hour
          sns.distplot(iphone_hour,hist=False,label='iPhone')
          sns.distplot(android_hour,hist=False,label='Android')
          plt.xlabel('hour')
          plt.ylabel('fraction')
          plt.title('Distribution of Tweet Hours for Different Tweet Souces')
          sns.set(font_scale=2)
```

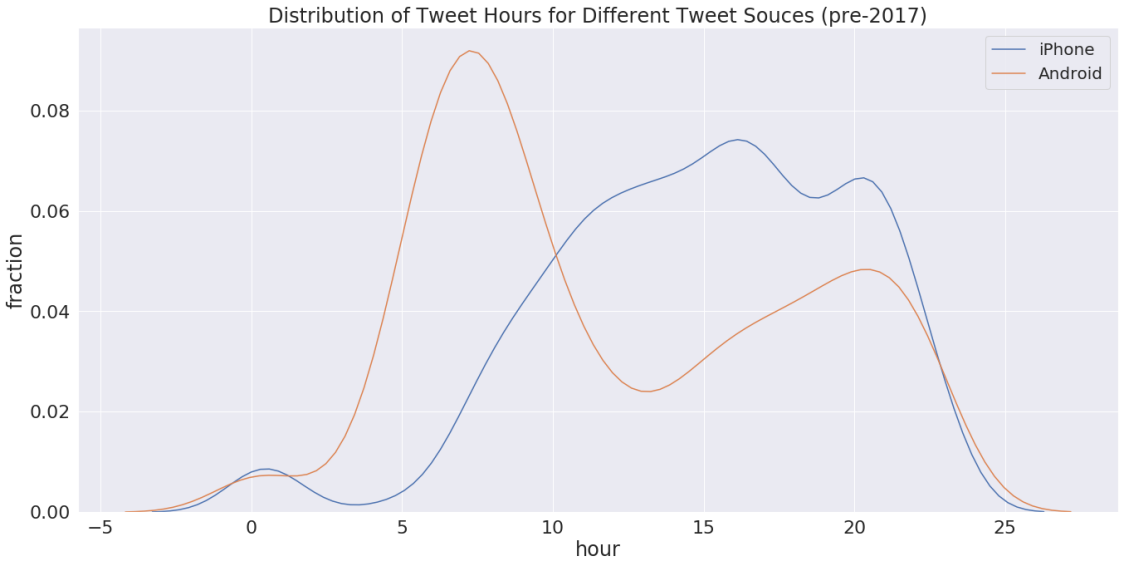


0.0.2 Question 4c

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 [460]: ### make your plot here
          bf2017_trump=trump[trump['year']<2017]
          iphone_hour_bf2017 = bf2017_trump[bf2017_trump['source']=='Twitter for iPhone']['est_time'].d
          android_hour_bf2017 = bf2017_trump[bf2017_trump['source']=='Twitter for Android']['est_time']
          sns.distplot(iphone_hour_bf2017,hist=False,label='iPhone')
          sns.distplot(android_hour_bf2017,hist=False,label='Android')
          plt.xlabel('hour')
          plt.ylabel('fraction')
          plt.title('Distribution of Tweet Hours for Different Tweet Souces (pre-2017)')
          plt.legend(prop={'size':20})
          sns.set(font_scale=2)
```



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?

I think the information from the figure and can't support to this theory. Here are the information we can get from the two figures: 1)Trump had tweets from both iPhone and Android before 2017, and the total time spending on Android was far more than the time on iPhone, which can tell that Trump usually wrote tweets on Android personally. 2)Trump stopped using Android after 2017. Then the total time spending on iPhone increased, which could conclude that Trump shifted his tweets from Android to iPhone after that, however, it was still much lower than the time spending on Android before. 3)Before 2017, he tweeted more on Android from 0~10, and he tweeted more on iPhone from 10~22, however, but this could only tell that he used his two phones at different times, but could not tell that his staff used his iPhone for tweeting. Additional analysis on the number of days that Donald Trump used his iPhone and Android to tweet at the same time during the day will be helpful to find how many tweets from iPhones were posted at the same time as the tweets from Android during a day. If the result of that is big enough, it might help 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 [429]: tidy_format
```

```
Out[429]:                       num            word
          id
          786204978629185536     0             pay
          ... Omitting 58 lines ...
          1052268011900555265   29          pompeo

          [259520 rows x 2 columns]
```