02-Pandas Data Visualization Exercise

October 28, 2019

___ # Pandas Data Visualization Exercise

This is just a quick exercise for you to review the various plots we showed earlier. Use **df3** to replicate the following plots.

```
In [1]: import pandas as pd
        import matplotlib.pyplot as plt
        df3 = pd.read_csv('df3')
        %matplotlib inline
```

```
In [2]: df3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 4 columns):
a    500 non-null float64
b    500 non-null float64
c    500 non-null float64
d    500 non-null float64
dtypes: float64(4)
memory usage: 15.8 KB
```

```
In [3]: df3.head()
```

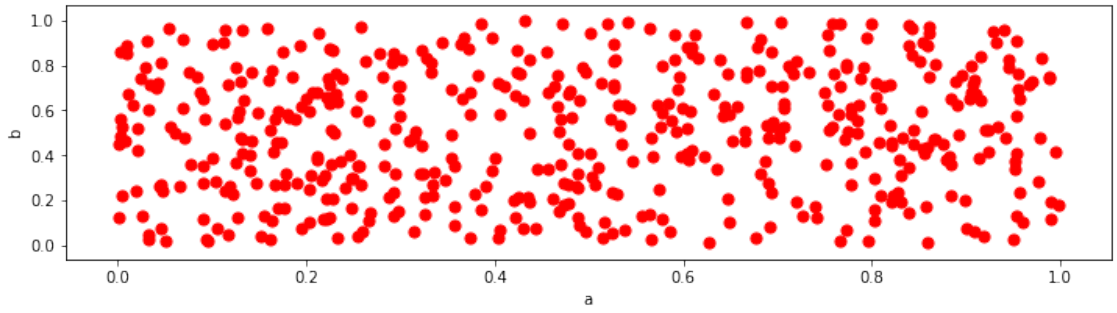
Out[3]:

	a	b	c	d
0	0.336272	0.325011	0.001020	0.401402
1	0.980265	0.831835	0.772288	0.076485
2	0.480387	0.686839	0.000575	0.746758
3	0.502106	0.305142	0.768608	0.654685
4	0.856602	0.171448	0.157971	0.321231

** Recreate this scatter plot of b vs a. Note the color and size of the points. Also note the figure size. See if you can figure out how to stretch it in a similar fashion. Remeber back to your matplotlib lecture...**

```
In [21]: df3.plot.scatter(x = 'a',y='b', c= 'red', figsize = (12,3),s=50)
```

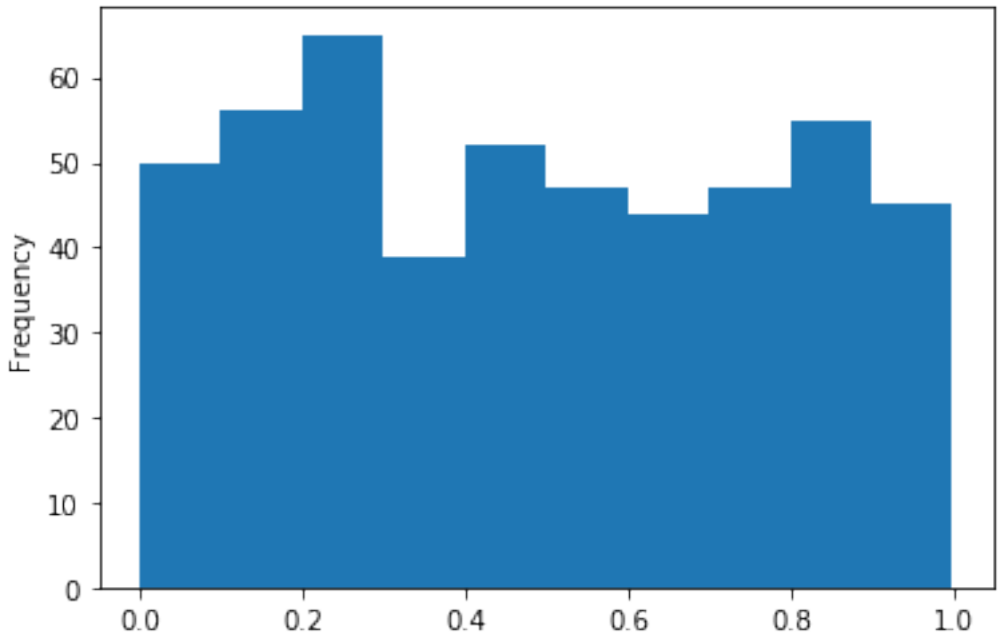
```
Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x120afad30>
```



** Create a histogram of the 'a' column.**

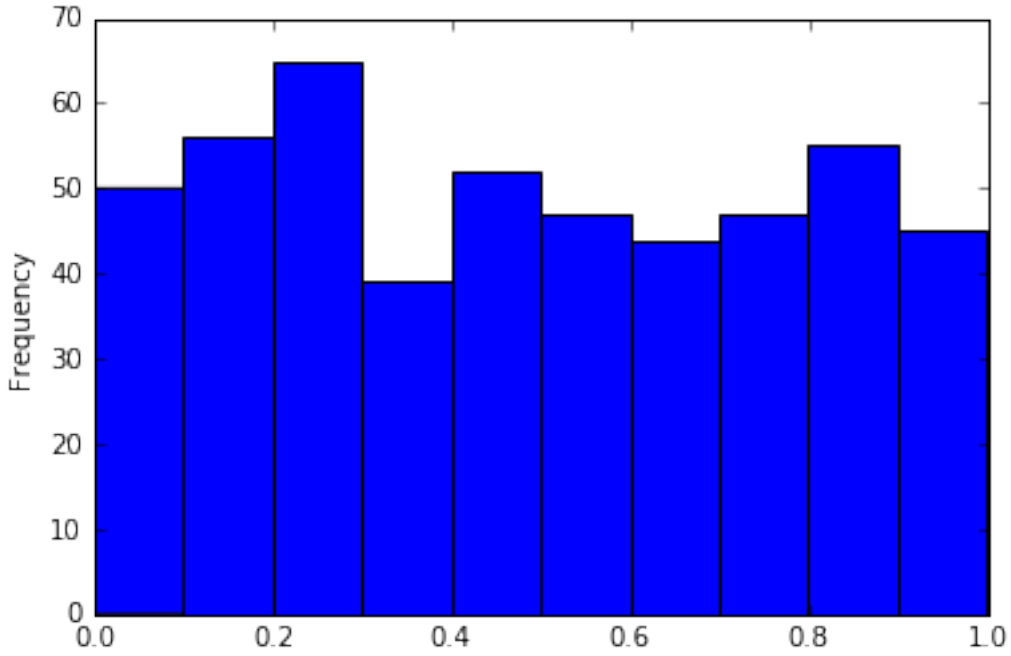
```
In [22]: df3['a'].plot.hist()
```

```
Out[22]: <matplotlib.axes._subplots.AxesSubplot at 0x120f03710>
```



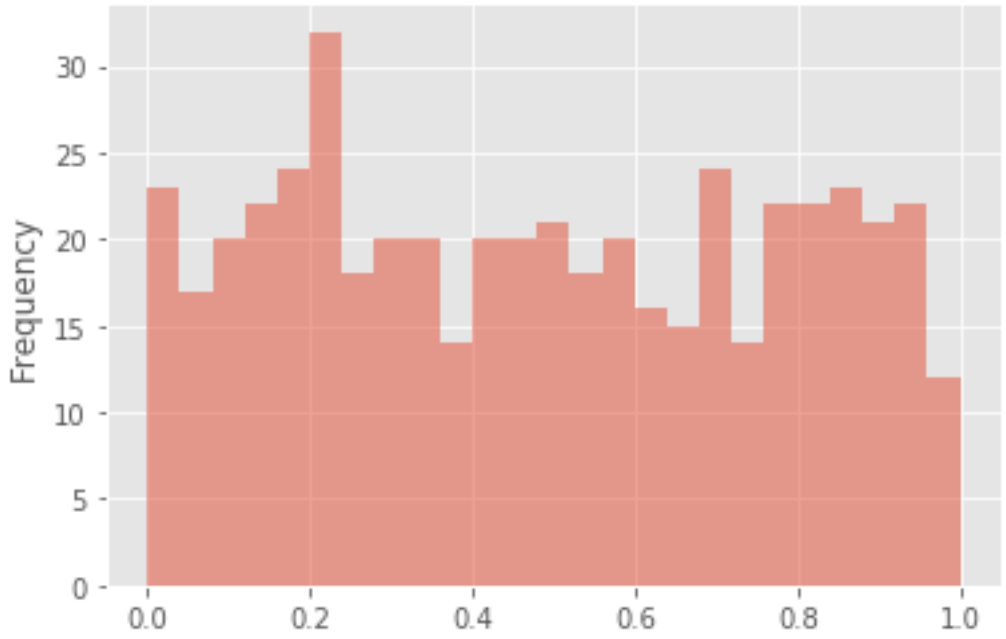

```
In [5]:
```

```
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x1177a2860>
```



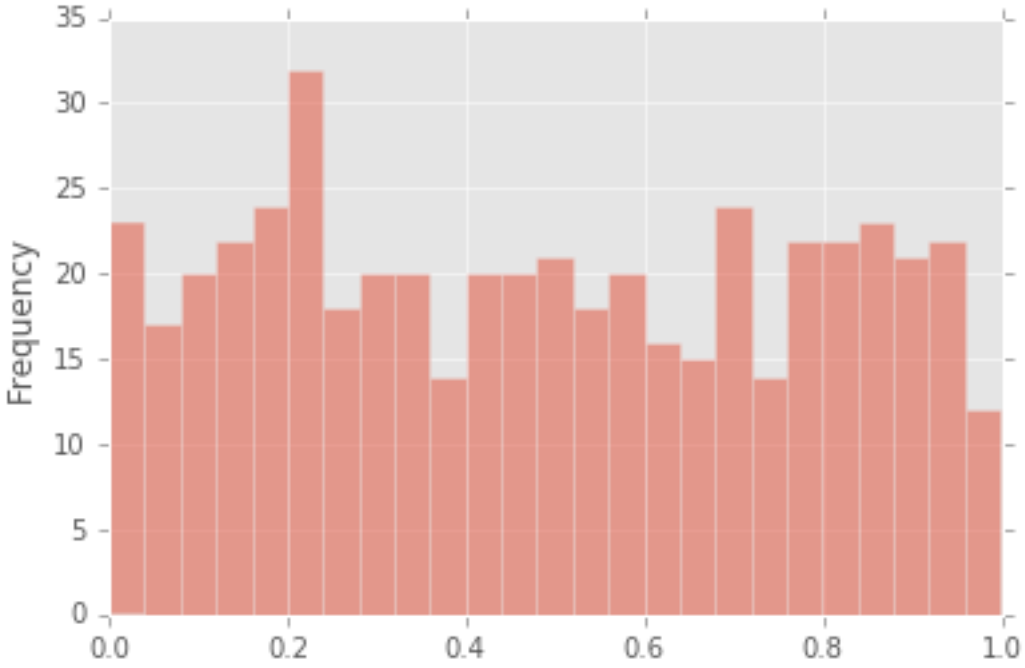
** These plots are okay, but they don't look very polished. Use style sheets to set the style to 'ggplot' and redo the histogram from above. Also figure out how to add more bins to it.***

```
In [26]: import matplotlib.pyplot as plt

In [27]: plt.style.use('ggplot')

In [31]: df3['a'].plot.hist(bins=25, alpha = 0.5)

Out[31]: <matplotlib.axes._subplots.AxesSubplot at 0x1a231858d0>
```



In [7]:

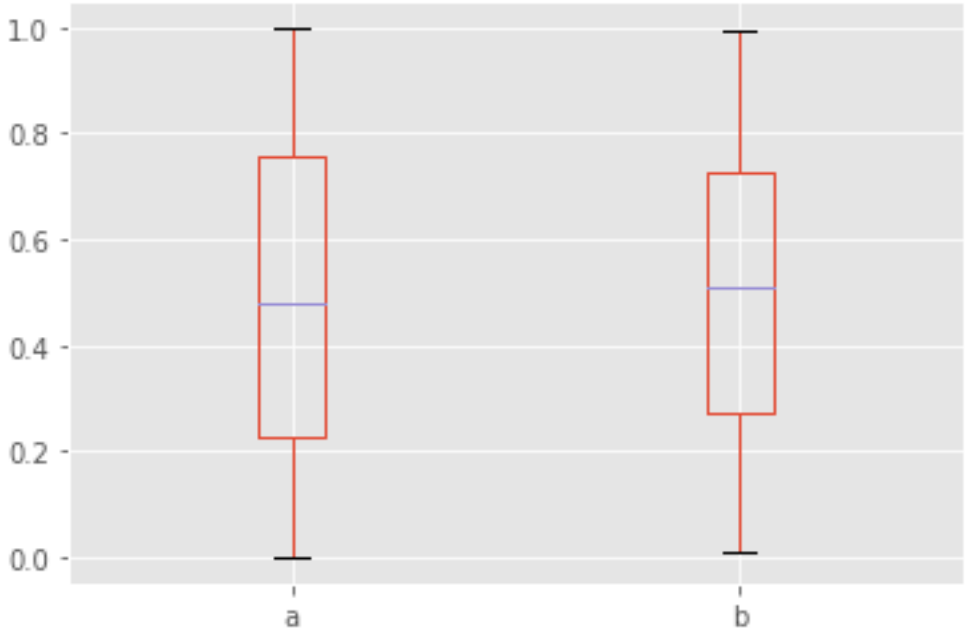
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x11a87b908>



** Create a boxplot comparing the a and b columns.**

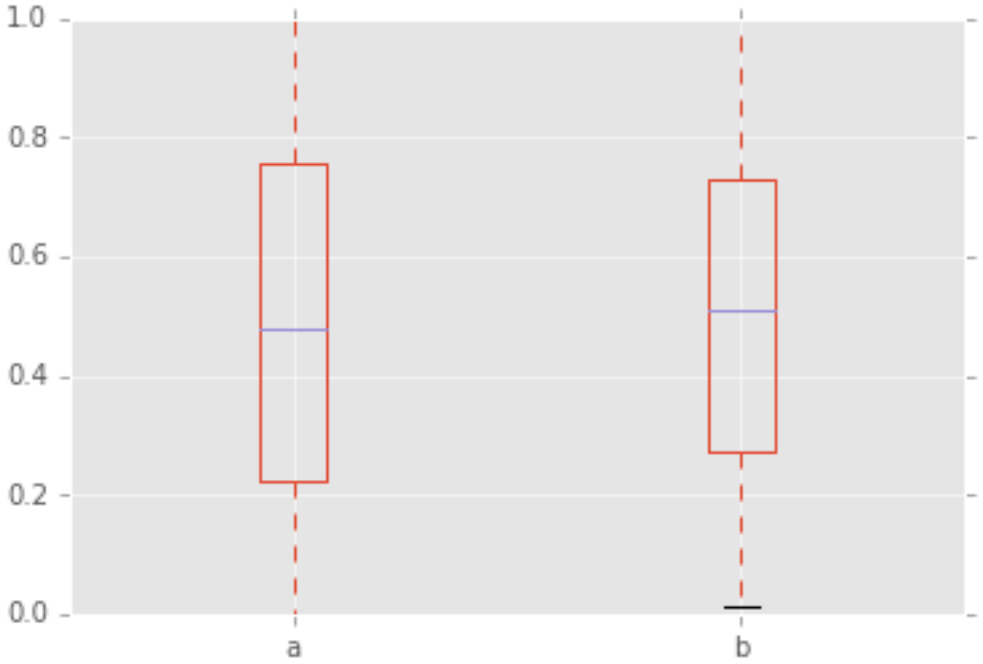
```
In [32]: df3[['a','b']].boxplot()
```

```
Out[32]: <matplotlib.axes._subplots.AxesSubplot at 0x1a231854e0>
```




```
In [8]:
```

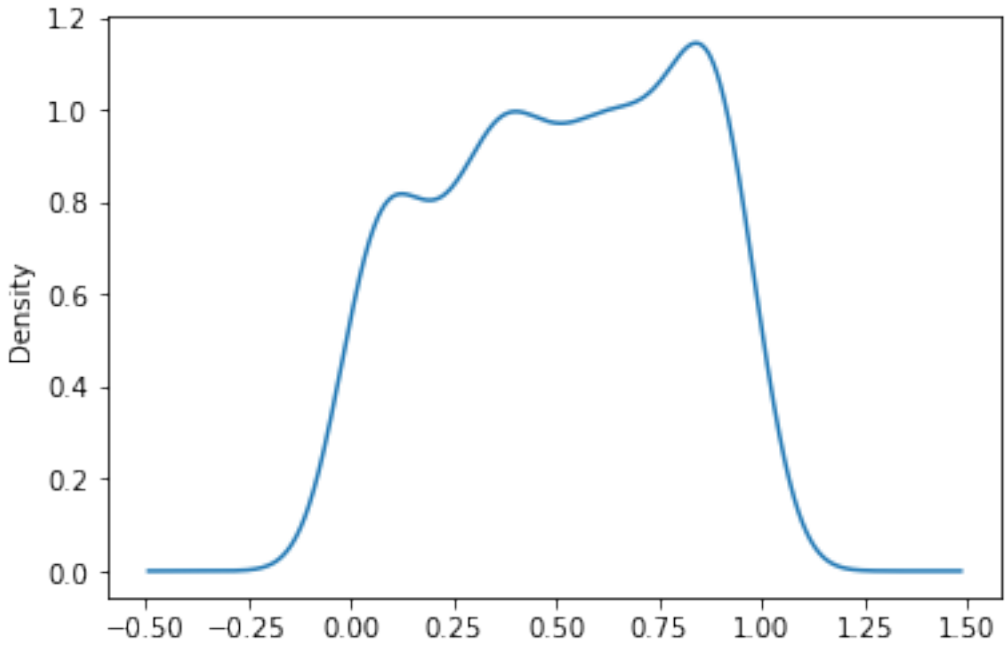
```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x1177c4a20>
```



** Create a kde plot of the 'd' column **

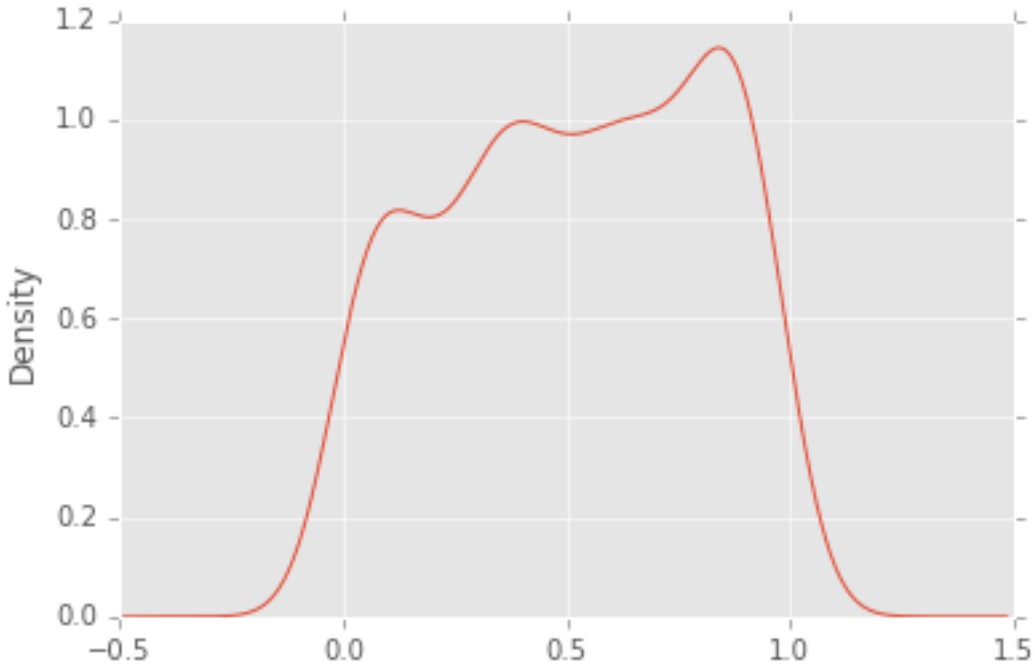
```
In [23]: df3['d'].plot.density()
```

```
Out[23]: <matplotlib.axes._subplots.AxesSubplot at 0x12101fa20>
```




```
In [9]:
```

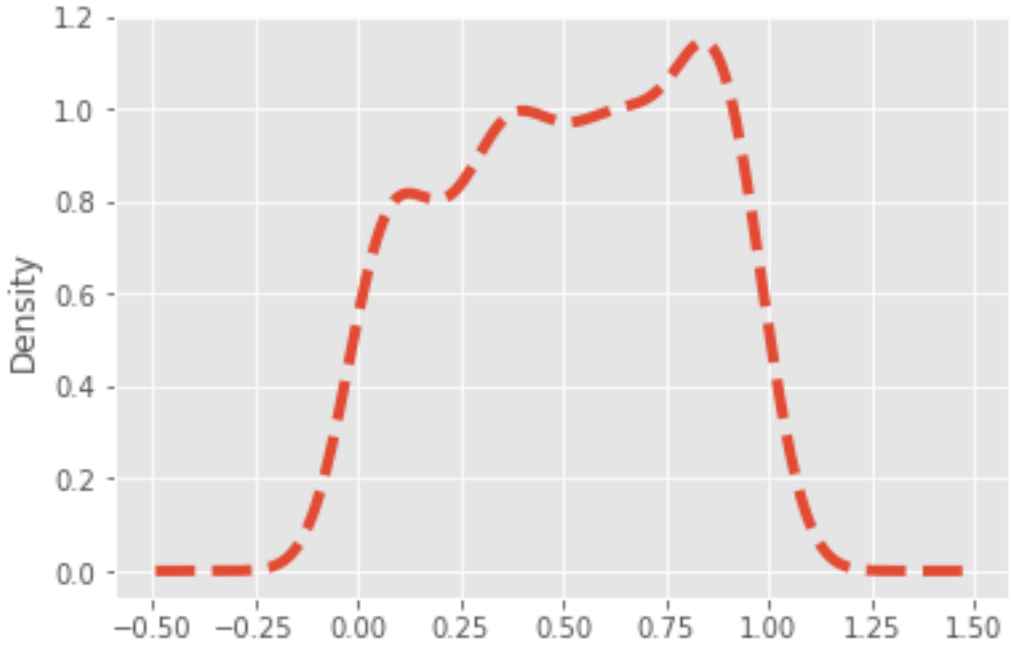
```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x11abb6278>
```



** Figure out how to increase the linewidth and make the linestyle dashed. (Note: You would usually not dash a kde plot line)**

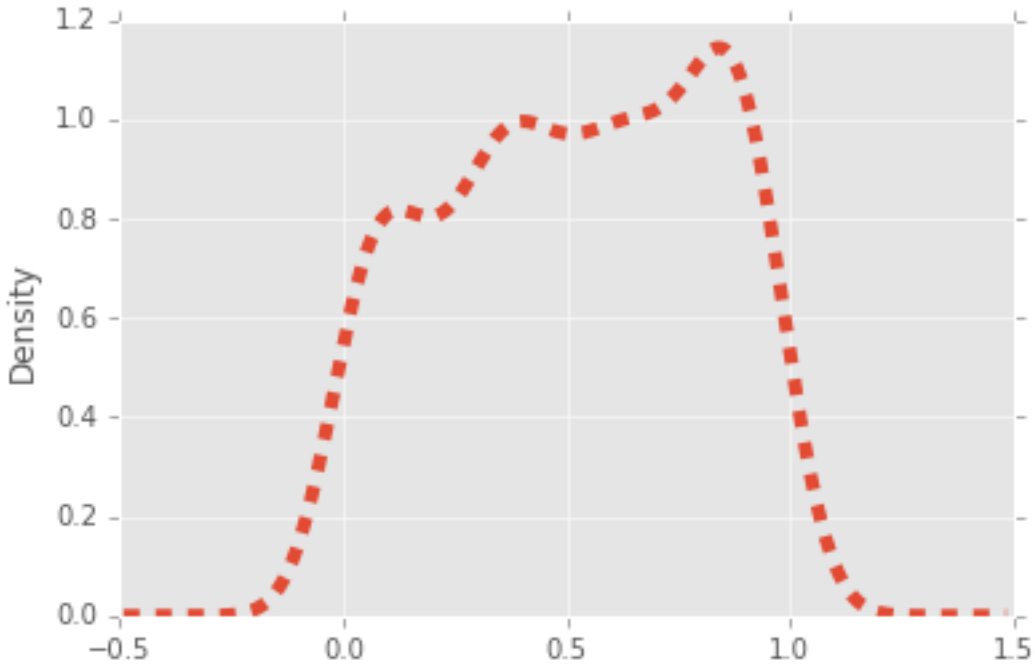
```
In [38]: df3['d'].plot.density(lw =4, ls = '--')
```

```
Out[38]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2353ac18>
```



In [10]:

Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x11ab9acc0>

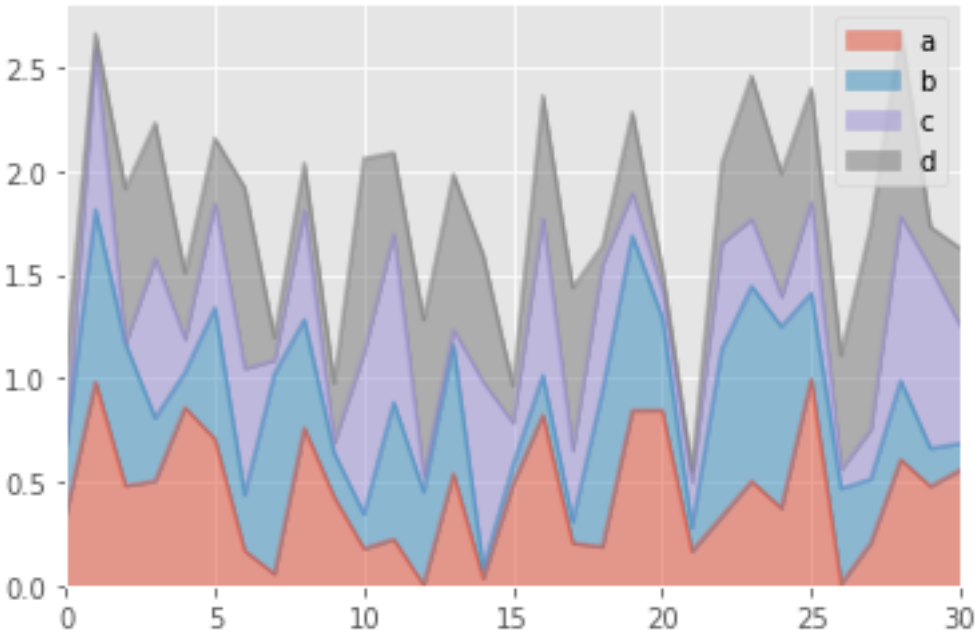


** Create an area plot of all the columns for just the rows up to 30. (hint: use .ix).**

In [43]:
```
df3.ix[0:30].plot.area(stacked=True,alpha=0.5)
```

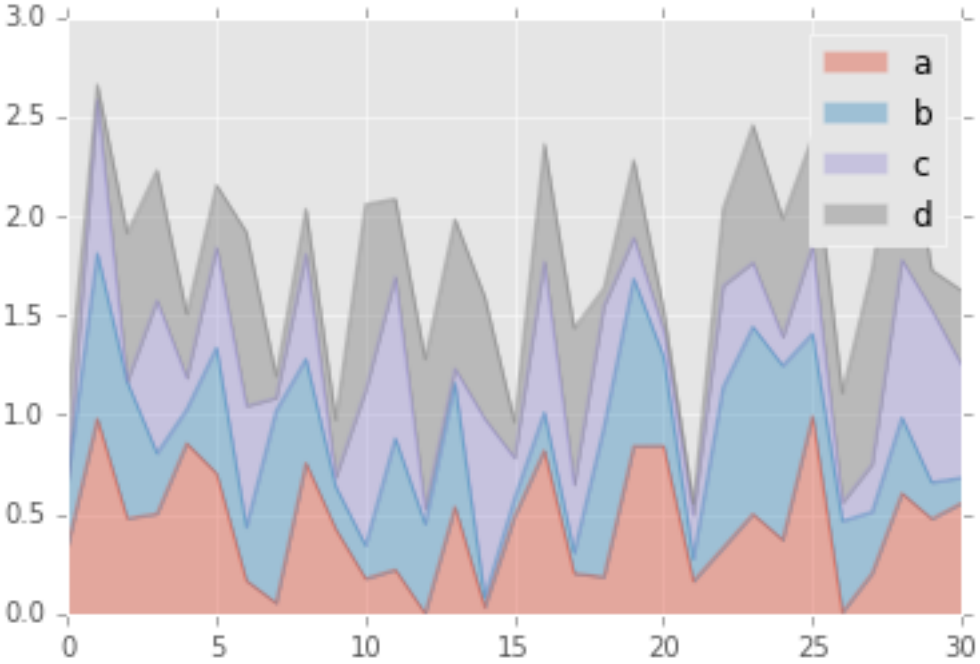
```
/Users/fernandomarin/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:1: FutureWarni
.ix is deprecated. Please use
.loc for label based indexing or
.iloc for positional indexing

See the documentation here:
http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#ix-indexer-is-deprecated
  """Entry point for launching an IPython kernel.
```

Out[43]: <matplotlib.axes._subplots.AxesSubplot at 0x1a23ba38d0>



In [15]:

Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x11ccdfbe0>



0.1 Bonus Challenge!

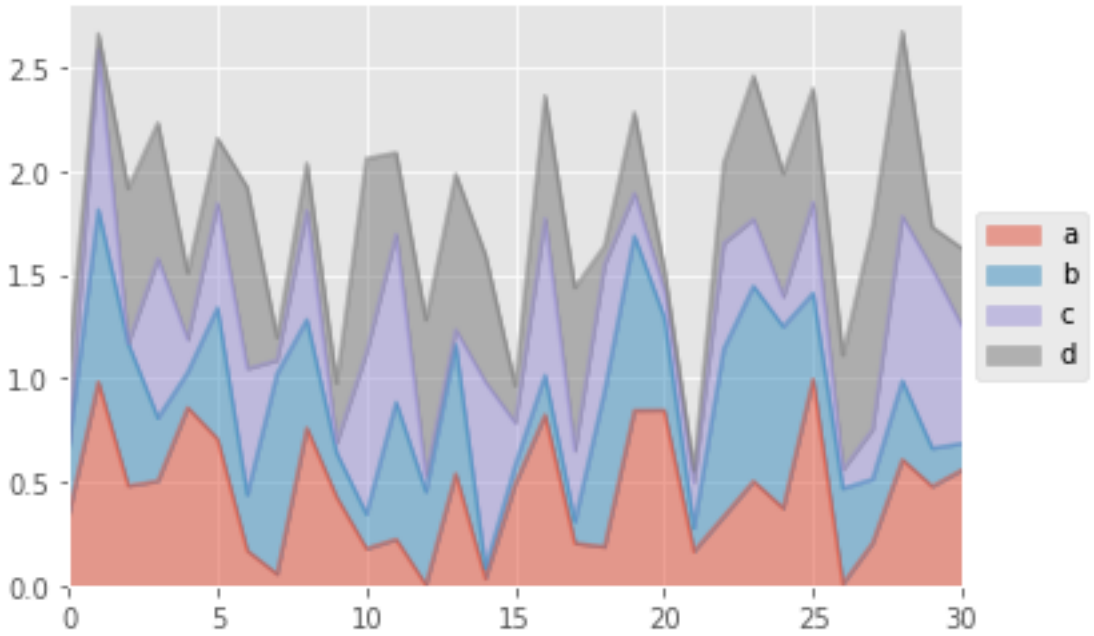
Note, you may find this really hard, reference the solutions if you can't figure it out! ** Notice how the legend in our previous figure overlapped some of actual diagram. Can you figure out how to display the legend outside of the plot as shown below?**

** Try searching Google for a good stackoverflow link on this topic. If you can't find it on your own - use this one for a hint.**

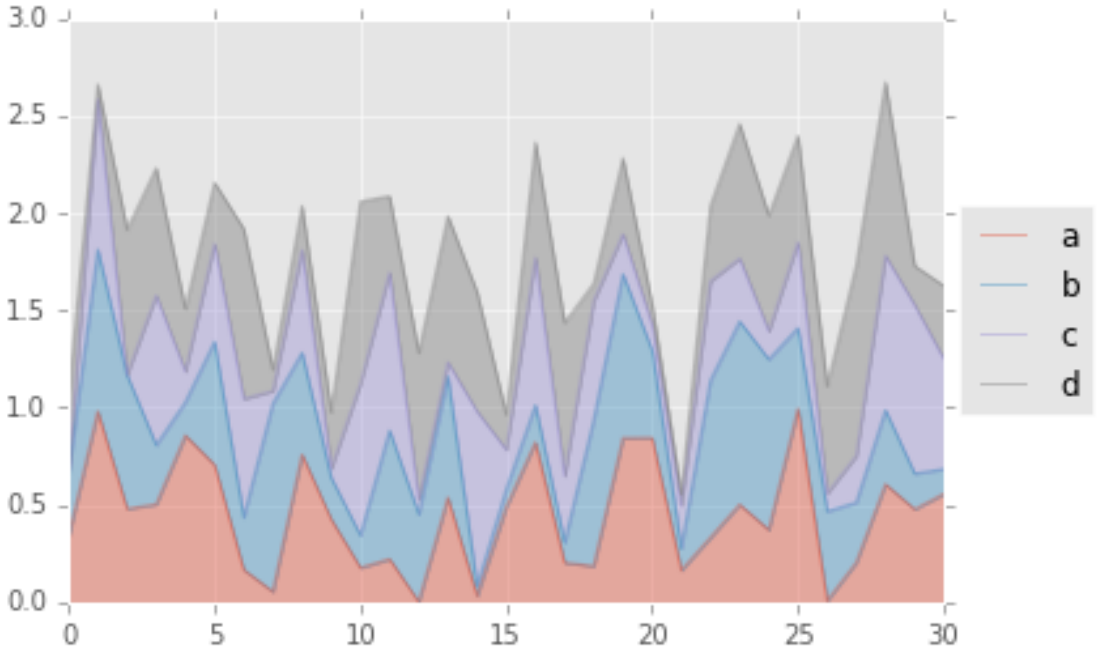
```
In [45]: df3.ix[0:30].plot.area(stacked=True,alpha=0.5)
         plt.legend(loc='center left', bbox_to_anchor = (1.0,0.5))
```

```
/Users/fernandomarin/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:1: FutureWarni
.ix is deprecated. Please use
.loc for label based indexing or
.iloc for positional indexing

See the documentation here:
http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#ix-indexer-is-deprecated
  """Entry point for launching an IPython kernel.
```

```
Out[45]: <matplotlib.legend.Legend at 0x1206bd668>
```




```
In [17]:
```



1 Great Job!