Rich Gruss

Python for Data Analysis, Ch. 7 and 8

Chapter 7

1. DataFrames can be merged on key values, much like an SQL join. Pandas even borrows some terminology from SQL, including "left," "right," "inner," and "outer." They can also be concatenated along an axis.
2. Stack() transforms tabular data into hierarchically indexed data. Unstack() transforms it back.

```
In [8]: %run /Users/rgruss/git/CMDA/notes/mypandas3.py
number    one  two  three
state
Ohio        0    1      2
Colorado    3    4      5

In [9]: %run /Users/rgruss/git/CMDA/notes/mypandas3.py
state     number
Ohio      one       0
          two       1
          three     2
Colorado  one       3
          two       4
          three     5
dtype: int64
```

3. To take a random sample of rows from your data set:
    a. get a permutation of the number of rows you want with "samples = np.random.permutation(numrows)"
    b. call df.take(samples)

Chapter 8

1. Matplotlib combined with an iPython notebook provides a MATLAB-like environment for visualizations.
2. Plots are contained within a Figure, and after subplots are added, you show the plots with figure.show()
3. Matplotlib includes tools for controlling ticks, adding legends, and drawing geometrical figures.
4. Other important toolkits for plotting include Chaco and Mayavi, but the future of visualization appears to be web-based JavaScript.