Rich Gruss

Python for Data Analysis, Chapter 5 and 6

Chapter 5:

1. Pandas is built on top of numpy and provides an R-like DataFrame, a Series, reductions, and relational database-style joins and projections.
2. A DataFrame can be initialized with a dictionary or with JSON, which has an identical object syntax.
3. A DataFrame supports hierarchical indexing
4. Important functions in pandas:
    a. DataFrame.ix[<position or name, or slice op>]: retrieves rows
    b. Numpy ufuncs (element-wise array operations) work on data frames.
    c. DataFrame.sum(),DataFrame.mean(), DataFrame.dropna()
    d. Series.value_counts() : gives distribution of values

Chapter 6

1. pandas provides two functions for easily reading data tables into a DataFrame: read_table() and read_csv().
2. You can create your own "dialect" of reader by subclassing csv.Dialect and setting values for lineterminator, delimiter, and quotechar.
3. Web pages can be easily downloaded and parsed with a handful of lines from the requests module or lxml.html and urllib2.