Pandas

Pandas is a package used for managing data.

Pandas main use is that it creates 2 new data types for storing data: series and dataframe.

Think of a pandas dataframe like an excel spreadsheet that is storing some data. One column can have customer name, one column can have product sold name, another column can have price or quantity... Then the rows could be individual sales.

A dataframe is made up of several series. Each column of a dataframe is a series.

We can name each column and row of a dataframe.

A pandas dataframe is very similar to a data.frame in R.

Similar to numpy arrays, a dataframe is a more robust data type for storing data than lists of lists. Dataframes are more flexible than numpy arrays.

A numpy array can create a matrix with all entries of the same data type. In a dataframe each column can have its own datatype.

That's not to say numpy arrays aren't useful. It is often easiest to convert some subset of a dataframe to a numpy array and then use that to do some math.

Pandas also has SQL-like functions for merging, joining, and sorting dataframes.

In [ ]:

```
import pandas as pd
import numpy as np  # numpy is not necessary for pandas, but we will
# in general it's good practice to import all pacakages at the beginr
```

In [ ]:

```
# first let's look at series - think of this as a single column of a
# each entry in a series corresponds to an individual row in the spre
# we can create a series by converting a list, or numpy array

mylist = [5.4,6.1,1.7,99.8]
myarray = np.array(mylist)
```

In [ ]:

```
myseries1 = pd.Series(data=mylist)
print(myseries1)
myseries2 = pd.Series(data=myarray)
print(myseries2)
```