

In [1]:

```
import numpy as np
from pandas import Series, DataFrame
import pandas as pd
```

In [2]:

```
#Now we'll learn DataFrames

#Let's get some data to play with. How about the NFL?
import webbrowser
website = 'http://en.wikipedia.org/wiki/NFL_win-loss_records'
webbrowser.open(website)
```

Out[2]:

True

In [3]:

```
#Copy and read to get data
nfl_frame = pd.read_clipboard()
```

In [4]:

```
#Show
nfl_frame
```

Out[4]:

	Rank	Team	GP	Won	Lost	Tied	Pct.	First NFL Season	Division
0	1	Dallas Cowboys	914	520	388	6	0.572	1960	NFC East
1	2	Green Bay Packers	1368	756	574	38	0.567	1921	NFC North
2	3	New England Patriots	916	512	395	9	0.564	1960	AFC East
3	4	Chicago Bears	1402	769	591	42	0.563	1920	NFC North
4	5	Baltimore Ravens	384	214	169	1	0.559	1996	AFC North
5	6	Miami Dolphins	832	457	371	4	0.552	1966	AFC East
6	7	Minnesota Vikings	902	488	403	11	0.547	1961	NFC North
7	8	San Francisco 49ers	1034	545	475	14	0.534	1950	NFC West
8	9	New York Giants	1,337	696	608	33	0.533	1925	NFC East
9	10	Denver Broncos	916	483	423	10	0.533	1960	AFC West

In [5]:

```
# We can grab the column names with .columns
nfl_frame.columns
```

Out[5]:

```
Index(['Rank', 'Team', 'GP', 'Won', 'Lost', 'Tied', 'Pct.', 'First NFL Sea-
son',
      'Division'],
      dtype='object')
```

In [6]:

```
nfl_frame["Team"]
```

Out[6]:

```
0      Dallas Cowboys
1    Green Bay Packers
2  New England Patriots
3      Chicago Bears
4    Baltimore Ravens
5      Miami Dolphins
6    Minnesota Vikings
7    San Francisco 49ers
8      New York Giants
9      Denver Broncos
Name: Team, dtype: object
```

In [7]:

```
nfl_frame.head(5)
```

Out[7]:

	Rank	Team	GP	Won	Lost	Tied	Pct.	First NFL Season	Division
0	1	Dallas Cowboys	914	520	388	6	0.572	1960	NFC East
1	2	Green Bay Packers	1368	756	574	38	0.567	1921	NFC North
2	3	New England Patriots	916	512	395	9	0.564	1960	AFC East
3	4	Chicago Bears	1402	769	591	42	0.563	1920	NFC North
4	5	Baltimore Ravens	384	214	169	1	0.559	1996	AFC North

In [8]:

```
#Lets see some specific data columns
DataFrame(nfl_frame,columns=['Team', 'First NFL Season', 'Total Games', 'GP'])
```

Out[8]:

	Team	First NFL Season	Total Games	GP
0	Dallas Cowboys	1960	NaN	914
1	Green Bay Packers	1921	NaN	1368
2	New England Patriots	1960	NaN	916
3	Chicago Bears	1920	NaN	1402
4	Baltimore Ravens	1996	NaN	384
5	Miami Dolphins	1966	NaN	832
6	Minnesota Vikings	1961	NaN	902
7	San Francisco 49ers	1950	NaN	1034
8	New York Giants	1925	NaN	1,337
9	Denver Broncos	1960	NaN	916

In [9]:

```
#What happens if we ask for a column that doesn't exist?
DataFrame(nfl_frame,columns=['Team', 'First NFL Season', 'GP', 'Division'])
```

Out[9]:

	Team	First NFL Season	GP	Division
0	Dallas Cowboys	1960	914	NFC East
1	Green Bay Packers	1921	1368	NFC North
2	New England Patriots	1960	916	AFC East
3	Chicago Bears	1920	1402	NFC North
4	Baltimore Ravens	1996	384	AFC North
5	Miami Dolphins	1966	832	AFC East
6	Minnesota Vikings	1961	902	NFC North
7	San Francisco 49ers	1950	1034	NFC West
8	New York Giants	1925	1,337	NFC East
9	Denver Broncos	1960	916	AFC West

In [10]:

```
# Call columns  
nfl_frame.columns
```

Out[10]:

```
Index(['Rank', 'Team', 'GP', 'Won', 'Lost', 'Tied', 'Pct.', 'First NFL Sea  
son',  
      'Division'],  
      dtype='object')
```

In [11]:

```
#We can retrieve individual columns  
nfl_frame['Team']
```

Out[11]:

```
0      Dallas Cowboys  
1    Green Bay Packers  
2  New England Patriots  
3    Chicago Bears  
4    Baltimore Ravens  
5    Miami Dolphins  
6    Minnesota Vikings  
7    San Francisco 49ers  
8    New York Giants  
9    Denver Broncos  
Name: Team, dtype: object
```

In [12]:

```
# Or try this method for multiple word columns  
nfl_frame['GP']
```

Out[12]:

```
0      914  
1    1368  
2      916  
3    1402  
4      384  
5      832  
6      902  
7    1034  
8    1,337  
9      916  
Name: GP, dtype: object
```

In [13]:

```
#We can retrieve rows through indexing
nfl_frame.iloc[3]
```

Out[13]:

```
Rank          4
Team          Chicago Bears
GP           1402
Won           769
Lost          591
Tied           42
Pct.          0.563
First NFL Season 1920
Division      NFC North
Name: 3, dtype: object
```

In [14]:

```
#We can also assign value to entire columns
nfl_frame['Stadium']="Levi's Stadium" #Careful with the ' here
```

In [15]:

```
nfl_frame
```

Out[15]:

	Rank	Team	GP	Won	Lost	Tied	Pct.	First NFL Season	Division	Stadium
0	1	Dallas Cowboys	914	520	388	6	0.572	1960	NFC East	Levi's Stadium
1	2	Green Bay Packers	1368	756	574	38	0.567	1921	NFC North	Levi's Stadium
2	3	New England Patriots	916	512	395	9	0.564	1960	AFC East	Levi's Stadium
3	4	Chicago Bears	1402	769	591	42	0.563	1920	NFC North	Levi's Stadium
4	5	Baltimore Ravens	384	214	169	1	0.559	1996	AFC North	Levi's Stadium
5	6	Miami Dolphins	832	457	371	4	0.552	1966	AFC East	Levi's Stadium
6	7	Minnesota Vikings	902	488	403	11	0.547	1961	NFC North	Levi's Stadium
7	8	San Francisco 49ers	1034	545	475	14	0.534	1950	NFC West	Levi's Stadium
8	9	New York Giants	1,337	696	608	33	0.533	1925	NFC East	Levi's Stadium
9	10	Denver Broncos	916	483	423	10	0.533	1960	AFC West	Levi's Stadium

In [16]:

```
#Putting numbers for stadiums
nfl_frame["Stadium"] = np.arange(10)

#Show
nfl_frame
```

Out[16]:

	Rank	Team	GP	Won	Lost	Tied	Pct.	First NFL Season	Division	Stadium
0	1	Dallas Cowboys	914	520	388	6	0.572	1960	NFC East	0
1	2	Green Bay Packers	1368	756	574	38	0.567	1921	NFC North	1
2	3	New England Patriots	916	512	395	9	0.564	1960	AFC East	2
3	4	Chicago Bears	1402	769	591	42	0.563	1920	NFC North	3
4	5	Baltimore Ravens	384	214	169	1	0.559	1996	AFC North	4
5	6	Miami Dolphins	832	457	371	4	0.552	1966	AFC East	5
6	7	Minnesota Vikings	902	488	403	11	0.547	1961	NFC North	6
7	8	San Francisco 49ers	1034	545	475	14	0.534	1950	NFC West	7
8	9	New York Giants	1,337	696	608	33	0.533	1925	NFC East	8
9	10	Denver Broncos	916	483	423	10	0.533	1960	AFC West	9

In [17]:

```
# Call columns
nfl_frame.columns
```

Out[17]:

```
Index(['Rank', 'Team', 'GP', 'Won', 'Lost', 'Tied', 'Pct.', 'First NFL Sea
son',
      'Division', 'Stadium'],
      dtype='object')
```

In [18]:

```
#Adding a Series to a DataFrame
stadiums = Series(["Levi's Stadium", "AT&T Stadium"], index=[4,0])
```

In [19]:

```
#Now input into the nfl DataFrame
nfl_frame['Stadium']=stadiums

#Show
nfl_frame
```

Out[19]:

	Rank	Team	GP	Won	Lost	Tied	Pct.	First NFL Season	Division	Stadium
0	1	Dallas Cowboys	914	520	388	6	0.572	1960	NFC East	AT&T Stadium
1	2	Green Bay Packers	1368	756	574	38	0.567	1921	NFC North	NaN
2	3	New England Patriots	916	512	395	9	0.564	1960	AFC East	NaN
3	4	Chicago Bears	1402	769	591	42	0.563	1920	NFC North	NaN
4	5	Baltimore Ravens	384	214	169	1	0.559	1996	AFC North	Levi's Stadium
5	6	Miami Dolphins	832	457	371	4	0.552	1966	AFC East	NaN
6	7	Minnesota Vikings	902	488	403	11	0.547	1961	NFC North	NaN
7	8	San Francisco 49ers	1034	545	475	14	0.534	1950	NFC West	NaN
8	9	New York Giants	1,337	696	608	33	0.533	1925	NFC East	NaN
9	10	Denver Broncos	916	483	423	10	0.533	1960	AFC West	NaN

In [20]:

```
#We can also delete columns
del nfl_frame['Stadium']

nfl_frame
```

Out[20]:

	Rank	Team	GP	Won	Lost	Tied	Pct.	First NFL Season	Division
0	1	Dallas Cowboys	914	520	388	6	0.572	1960	NFC East
1	2	Green Bay Packers	1368	756	574	38	0.567	1921	NFC North
2	3	New England Patriots	916	512	395	9	0.564	1960	AFC East
3	4	Chicago Bears	1402	769	591	42	0.563	1920	NFC North
4	5	Baltimore Ravens	384	214	169	1	0.559	1996	AFC North
5	6	Miami Dolphins	832	457	371	4	0.552	1966	AFC East
6	7	Minnesota Vikings	902	488	403	11	0.547	1961	NFC North
7	8	San Francisco 49ers	1034	545	475	14	0.534	1950	NFC West
8	9	New York Giants	1,337	696	608	33	0.533	1925	NFC East
9	10	Denver Broncos	916	483	423	10	0.533	1960	AFC West

In [21]:

```
#DataFrames can be constructed many ways. Another way is from a dictionary of equal length lists
data = {'City':['SF','LA','NYC'],
        'Population':[837000,3880000,8400000]}

city_frame = DataFrame(data)

#Show
city_frame
```

Out[21]:

	City	Population
0	SF	837000
1	LA	3880000
2	NYC	8400000

In [22]:

```
#For full list of ways to create DataFrames from various sources go to the documentation for pandas:  
website = 'https://pandas.pydata.org/pandas-docs/dev/reference/api/pandas.DataFrame.html'  
webbrowser.open(website)
```

Out[22]:

True

In []: