Intro to Programming for Public Policy Week 6 Pandas Addendum

Eric Potash

May 1, 2018

Writing/Reading CSVs

to_csv()

Use DataFrame.to_csv() to write a DataFrame to CSV:

```
>>> state_years
   pop    state  year
0  1.5     Ohio  2000
1  1.7     Ohio  2001
2  3.6     Ohio  2002
3  2.4   Nevada  2001
4  2.9   Nevada  2002

>>> state_years.to_csv('state_years.csv')
```

CSV contents

By default, index is included in output:

```
$ cat state_years.csv
,pop,state,year
0,1.5,Ohio,2000
1,1.7,Ohio,2001
2,3.6,Ohio,2002
3,2.4,Nevada,2001
4,2.9,Nevada,2002
```

read_csv()

When read back in with `read_csv()` this is annoying:

```
>>> pd.read_csv('state_years.csv')
   Unnamed: 0  pop    state  year
0           0  1.5     Ohio  2000
1           1  1.7     Ohio  2001
2           2  3.6     Ohio  2002
3           3  2.4   Nevada  2001
4           4  2.9   Nevada  2002
```

Solution

To avoid this, pass index=False to `to_csv()`:

```
>>> state_years.to_csv('state_years.csv', index=False)
```

```
$ cat state_years.csv
pop,state,year
1.5,Ohio,2000
1.7,Ohio,2001
3.6,Ohio,2002
2.4,Nevada,2001
2.9,Nevada,2002
```