
apricot_data

Unknown Author

January 09, 2014

Part I

IPython with bash

IPython can interact with other languages besides python. In the code below, a simple bash script is run to download and clean up UN data on apricot production.

```
In [1]: %%bash
curl > "../input/apricot_data.zip" "http://data.un.org/Handlers/DownloadHandler.ashx?D
unzip -p ../input/apricot_data.zip > ../input/apricot_data.csv
sed -i -r 's/([a-z]),/\1/' ../input/apricot_data.csv #remove delimiter from country n
grep '+' ../input/apricot_data.csv > ../input/apricot_regions.csv
grep -v '+' ../input/apricot_data.csv > ../input/apricot_countries.csv

% Total      % Received % Xferd  Average Speed   Time    Time     Time
Current

                        Dload  Upload   Total   Spent    Left
Speed
    0      0    0    0    0    0    0    0  --:--:--  --:--:--
--:--:--      0    0    0    0    0    0    0    0  --:--:--
0:00:01 --:--:--      0    0    0    0    0    0    0    0
--:--:--  0:00:02 --:--:--      0    0    0    0    0    0    0
0      0  --:--:--  0:00:03 --:--:--      0 100 68264  100 68264    0
0 19278      0 0:00:03 0:00:03 --:--:-- 19278
```

Part II

Using pandas to quickly inspect a data set

```
In [3]: %matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt
```

Read in the data:

```
In [4]: data = pd.read_csv('../input/apricot_countries.csv')
data.head()
```

	Country or Area	Element Code	Element	Year	Unit	Value	\
Out [4]: 0	Afghanistan	31	Area Harvested	2007	Ha	3400	
1	Afghanistan	31	Area Harvested	2006	Ha	8030	
2	Afghanistan	31	Area Harvested	2005	Ha	5200	
3	Afghanistan	31	Area Harvested	2004	Ha	5200	
4	Afghanistan	31	Area Harvested	2003	Ha	7007	

	Value	Footnotes
0		F
1		NaN
2		F
3		F
4		NaN

Reshape the data to have year as the index, country as the columns, and

area harvested as the value:

```
In [5]: ahdata = data[data['Element'] == 'Area Harvested']
ahdata = ahdata.pivot(index='Year', columns='Country or Area', values='Value')
ahdata.ix[:, :10].head()
```

Country or Area	Afghanistan	Albania	Algeria	Argentina	Armenia
Out [5]: Australia \					
Year					
1961	4820	0	4200	0	NaN
3794					
1962	4820	0	4600	0	NaN
3794					
1963	4820	0	4000	0	NaN
4005					
1964	5100	0	4200	0	NaN
4059					
1965	5370	0	4000	0	NaN
3790					

Country or Area	Austria	Azerbaijan	Bosnia and Herzegovina	Bulgaria
Year				
1961	0	NaN		NaN
1962	0	NaN		NaN
1963	0	NaN		NaN
1964	0	NaN		NaN
1965	0	NaN		NaN

Plot area harvested for each country:

```
In [6]: for col in ahdata.columns:
plt.figure(figsize=[8,3])
plt.title(col)
plt.ylabel('Area Harvested (Ha)')
years = [int(y) for y in ahdata.index.values]
plt.plot(years, ahdata.ix[:, col])
plt.xticks(years, years, rotation=90, size='small')
```

c:\Users\Frank\Anaconda\lib\site-packages\matplotlib\pyplot.py:412:
RuntimeWarning: More than 20 figures have been opened. Figures created
through the pyplot interface ('matplotlib.pyplot.figure') are retained
until explicitly closed and may consume too much memory. (To control

this warning, see the rcParam 'figure.max_num_figures').
max_open_warning, RuntimeError)















































