[(http://www.pieriandata.com)](http://www.pieriandata.com)

Copyright Pierian Data
For more information, visit us at [www.pieriandata.com (http://www.pieriandata.com)](http://www.pieriandata.com)

Pandas Exercises

Time to test your new pandas skills! Use the `population_by_county.csv` file in the Data folder to complete the tasks in bold below!

> **NOTE:** ALL TASKS CAN BE DONE IN ONE LINE OF PANDAS CODE. GET STUCK? NO PROBLEM! CHECK OUT THE SOLUTIONS LECTURE!

> **IMPORTANT NOTE!** Make sure you don't run the cells directly above the example output shown, otherwise you will end up writing over the example output!

1. Import pandas and read in the `population_by_county.csv` file into a dataframe called `pop`.

In [1]:

```python
import pandas as pd
```

In [5]:

```python
pop = pd.read_csv('population_by_county.csv')
pop
```

Out[5]:

	County	State	2010Census	2017PopEstimate
0	Abbeville County	South Carolina	25417	24722
1	Acadia Parish	Louisiana	61773	62590
2	Accomack County	Virginia	33164	32545
3	Ada County	Idaho	392365	456849
4	Adair County	Iowa	7682	7054
...	...	...	...	...
3137	Yuma County	Arizona	195751	207534
3138	Yuma County	Colorado	10043	10075
3139	Zapata County	Texas	14018	14322
3140	Zavala County	Texas	11677	11948
3141	Ziebach County	South Dakota	2801	2756

3142 rows × 4 columns

2. Show the head of the dataframe

In [7]:

```python
# CODE HERE
pop.head()
```

Out[7]:

	County	State	2010Census	2017PopEstimate
0	Abbeville County	South Carolina	25417	24722
1	Acadia Parish	Louisiana	61773	62590
2	Accomack County	Virginia	33164	32545
3	Ada County	Idaho	392365	456849
4	Adair County	Iowa	7682	7054

In [3]:

```
# DON'T WRITE HERE
```

Out[3]:

	County	State	2010Census	2017PopEstimate
0	Abbeville County	South Carolina	25417	24722
1	Acadia Parish	Louisiana	61773	62590
2	Accomack County	Virginia	33164	32545
3	Ada County	Idaho	392365	456849
4	Adair County	Iowa	7682	7054

3. What are the column names?

In [8]:

```
pop.columns
```

Out[8]:

```
Index(['County', 'State', '2010Census', '2017PopEstimate'], dtype='objec
t')
```

In [4]:

```
# DON'T WRITE HERE
```

Out[4]:

```
Index(['County', 'State', '2010Census', '2017PopEstimate'], dtype='objec
t')
```

4. How many States are represented in this data set? *Note: the data includes the District of Columbia*

In [9]:

```
pop.State.nunique()
```

Out[9]:

```
51
```

In [5]:

```
# DON'T WRITE HERE
```

Out[5]:

```
51
```

5. Get a list or array of all the states in the data set.

In [19]:

```
pd.unique(pop.State)
```

Out[19]:

```
array(['South Carolina', 'Louisiana', 'Virginia', 'Idaho', 'Iowa',
       'Kentucky', 'Missouri', 'Oklahoma', 'Colorado', 'Illinois',
       'Indiana', 'Mississippi', 'Nebraska', 'North Dakota', 'Ohio',
       'Pennsylvania', 'Washington', 'Wisconsin', 'Vermont', 'Minnesota',
       'Florida', 'North Carolina', 'California', 'New York', 'Wyoming',
       'Michigan', 'Alaska', 'Maryland', 'Kansas', 'Tennessee', 'Texas',
       'Maine', 'Arizona', 'Georgia', 'Arkansas', 'New Jersey',
       'South Dakota', 'Alabama', 'Oregon', 'West Virginia',
       'Massachusetts', 'Utah', 'Montana', 'New Hampshire', 'New Mexico',
       'Rhode Island', 'Nevada', 'District of Columbia', 'Connecticut',
       'Hawaii', 'Delaware'], dtype=object)
```

In [6]:

```
# DON'T WRITE HERE
```

Out[6]:

```
array(['South Carolina', 'Louisiana', 'Virginia', 'Idaho', 'Iowa',
       'Kentucky', 'Missouri', 'Oklahoma', 'Colorado', 'Illinois',
       'Indiana', 'Mississippi', 'Nebraska', 'North Dakota', 'Ohio',
       'Pennsylvania', 'Washington', 'Wisconsin', 'Vermont', 'Minnesota',
       'Florida', 'North Carolina', 'California', 'New York', 'Wyoming',
       'Michigan', 'Alaska', 'Maryland', 'Kansas', 'Tennessee', 'Texas',
       'Maine', 'Arizona', 'Georgia', 'Arkansas', 'New Jersey',
       'South Dakota', 'Alabama', 'Oregon', 'West Virginia',
       'Massachusetts', 'Utah', 'Montana', 'New Hampshire', 'New Mexico',
       'Rhode Island', 'Nevada', 'District of Columbia', 'Connecticut',
       'Hawaii', 'Delaware'], dtype=object)
```

6. What are the five most common County names in the U.S.?

In [24]:

```
pop['County'].value_counts().head(5)
```

Out[24]:

```
Washington County    30
Jefferson County     25
Franklin County      24
Jackson County       23
Lincoln County       23
Name: County, dtype: int64
```

In [7]:

```
# DON'T WRITE HERE
```

Out[7]:

```
Washington County    30
Jefferson County     25
Franklin County      24
Jackson County       23
Lincoln County       23
Name: County, dtype: int64
```

7. What are the top 5 most populated Counties according to the 2010 Census?

In [ ]:

In [8]:

```
# DON'T WRITE HERE
```

Out[8]:

	County	State	2010Census	2017PopEstimate
1713	Los Angeles County	California	9818605	10163507
628	Cook County	Illinois	5194675	5211263
1209	Harris County	Texas	4092459	4652980
1784	Maricopa County	Arizona	3817117	4307033
2501	San Diego County	California	3095313	3337685

8. What are the top 5 most populated States according to the 2010 Census?

In [ ]:

In [9]:

```
# DON'T WRITE HERE
```

Out[9]:

State	2010Census	2017PopEstimate
California	37253956	39536653
Texas	25145561	28304596
New York	19378102	19849399
Florida	18801310	20984400
Illinois	12830632	12802023

9. How many Counties have 2010 populations greater than 1 million?

In [ ]:

In [10]:

```
# DON'T WRITE HERE
```

Out[10]:

39

10. How many Counties don't have the word "County" in their name?

In [ ]:

In [11]:

```
# DON'T WRITE HERE
```

Out[11]:

135

11. Add a column that calculates the percent change between the 2010 Census and the 2017 Population Estimate

In [ ]:

```
# CODE HERE

# USE THIS TO SHOW THE RESULT
pop.head()
```

In [12]:

```
# DON'T WRITE HERE
```

Out[12]:

	County	State	2010Census	2017PopEstimate	PercentChange
0	Abbeville County	South Carolina	25417	24722	-2.734390
1	Acadia Parish	Louisiana	61773	62590	1.322584
2	Accomack County	Virginia	33164	32545	-1.866482
3	Ada County	Idaho	392365	456849	16.434697
4	Adair County	Iowa	7682	7054	-8.174954

Bonus: What States have the highest estimated percent change between the 2010 Census and the 2017 Population Estimate?
This will take several lines of code, as it requires a recalculation of PercentChange.

In [ ]:

```
# CODE HERE
```

In [13]:

```
# DON'T WRITE HERE
```

Out[13]:

	2010Census	2017PopEstimate	PercentChange
State			
District of Columbia	601723	693972	15.330808
Texas	25145561	28304596	12.562993
North Dakota	672591	755393	12.310899
Utah	2763885	3101833	12.227282
Florida	18801310	20984400	11.611372

GREAT JOB!