New York State Industry

Cecilia Pershyn
Assignment 2 Data Configuration

```
In [4]: %matplotlib inline

        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns; sns.set()
```

```
In [5]: NYS_employment_2017_1 = pd.read_csv('NYS_Employment_2017_1.csv')
        NYS_employment_2017_2 = pd.read_csv('NYS_Employment_2017_2.csv')
        NYS_employment_2017_3 = pd.read_csv('NYS_Employment_2017_3.csv')
        NYS_employment_2017_4 = pd.read_csv('NYS_Employment_2017_4.csv')
```

```
In [6]: NYS_employment_2017_1.head()
```

Out[6]:

	area_fips	own_code	industry_code	agglvl_code	size_code	year	qtr	disclosure_cod
0	36000	0	10	50	0	2017	1	NaN
1	36000	1	10	51	0	2017	1	NaN
2	36000	1	101	52	0	2017	1	NaN
3	36000	1	1013	53	0	2017	1	NaN
4	36000	1	102	52	0	2017	1	NaN

5 rows × 42 columns

Create a column in the dataframe that summarizes the employment level for all 3 months in the quarter. The result is the total employment level for industries in the quarter.

```
In [7]: NYS_employment_2017_1['quarter1_emplvl'] = NYS_employment_2017_1['month1
        _emplvl'] + NYS_employment_2017_1['month2_emplvl'] + NYS_employment_2017
        _1['month3_emplvl']
        NYS_employment_2017_1.head()
```

Out[7]:

	area_fips	own_code	industry_code	agglvl_code	size_code	year	qtr	disclosure_coc
0	36000	0	10	50	0	2017	1	NaN
1	36000	1	10	51	0	2017	1	NaN
2	36000	1	101	52	0	2017	1	NaN
3	36000	1	1013	53	0	2017	1	NaN
4	36000	1	102	52	0	2017	1	NaN

5 rows × 43 columns

Create a new dataframe using only the columns that are relevant to study. In this case I chose to look at industry, quarterly contributions, total quarterly wages, and total quarterly employment level.

```
In [8]: NYS_employment_2017_1 = NYS_employment_2017_1[['year','qtr','industry_co
        de','qtrly_contributions','total_qtrly_wages','quarter1_emplvl']]
        NYS_employment_2017_1.head()
```

Out[8]:

	year	qtr	industry_code	qtrly_contributions	total_qtrly_wages	quarter1_emplvl
0	2017	1	10	1529251900	182980473877	27325310
1	2017	1	10	0	2222130245	351238
2	2017	1	101	0	142815	13
3	2017	1	1013	0	142815	13
4	2017	1	102	0	2221987430	351225

Rename the columns to indicate they represent values for quarter 1.

In [9]:
```
NYS_industry_2017_1 = NYS_employment_2017_1.rename(columns={'qtrly_contr
ibutions':'qtr1_contributions','total_qtrly_wages':'qtr1_wages'})
NYS_industry_2017_1.head()
```

Out[9]:

	year	qtr	industry_code	qtr1_contributions	qtr1_wages	quarter1_emplvl
0	2017	1	10	1529251900	182980473877	27325310
1	2017	1	10	0	2222130245	351238
2	2017	1	101	0	142815	13
3	2017	1	1013	0	142815	13
4	2017	1	102	0	2221987430	351225

Repeat the two steps above for quarters 2-4 of 2017.

In [10]:
```
NYS_employment_2017_2['quarter2_emplvl'] = NYS_employment_2017_2['month1
_emplvl'] + NYS_employment_2017_2['month2_emplvl'] + NYS_employment_2017
_2['month3_emplvl']
NYS_employment_2017_2.head()
```

Out[10]:

	area_fips	own_code	industry_code	agglvl_code	size_code	year	qtr	disclosure_cod
0	36000	0	10	50	0	2017	2	NaN
1	36000	1	10	51	0	2017	2	NaN
2	36000	1	101	52	0	2017	2	NaN
3	36000	1	1013	53	0	2017	2	NaN
4	36000	1	102	52	0	2017	2	NaN

5 rows × 43 columns

In [11]:
```
NYS_employment_2017_2 = NYS_employment_2017_2[['year','qtr','industry_co
de','qtrly_contributions','total_qtrly_wages','quarter2_emplvl']]
NYS_industry_2017_2 = NYS_employment_2017_2.rename(columns={'qtrly_contr
ibutions':'qtr2_contributions','total_qtrly_wages':'qtr2_wages'})
NYS_industry_2017_2.head()
```

Out[11]:

	year	qtr	industry_code	qtr2_contributions	qtr2_wages	quarter2_emplvl
0	2017	2	10	494482035	150417320920	27964380
1	2017	2	10	0	2217175239	349367
2	2017	2	101	0	97634	14
3	2017	2	1013	0	97634	14
4	2017	2	102	0	2217077605	349353

```
In [12]: NYS_employment_2017_3['quarter3_emplvl'] = NYS_employment_2017_3['month1
         _emplvl'] + NYS_employment_2017_3['month2_emplvl'] + NYS_employment_2017
         _3['month3_emplvl']
         NYS_employment_2017_3 = NYS_employment_2017_3[['year','qtr','industry_co
         de','qtrly_contributions','total_qtrly_wages','quarter3_emplvl']]
         NYS_industry_2017_3 = NYS_employment_2017_3.rename(columns={'qtrly_contr
         ibutions':'qtr3_contributions','total_qtrly_wages':'qtr3_wages'})
         NYS_industry_2017_3.head()
```

Out[12]:

	year	qtr	industry_code	qtr3_contributions	qtr3_wages	quarter3_emplvl
0	2017	4	10	267758192	175843272174	28347532
1	2017	4	10	0	2315826762	347186
2	2017	4	101	0	91119	10
3	2017	4	1013	0	91119	10
4	2017	4	102	0	2315735643	347176

```
In [13]: NYS_employment_2017_4['quarter4_emplvl'] = NYS_employment_2017_4['month1
         _emplvl'] + NYS_employment_2017_4['month2_emplvl'] + NYS_employment_2017
         _4['month3_emplvl']
         NYS_employment_2017_4 = NYS_employment_2017_4[['year','qtr','industry_co
         de','qtrly_contributions','total_qtrly_wages','quarter4_emplvl']]
         NYS_industry_2017_4 = NYS_employment_2017_4.rename(columns={'qtrly_contr
         ibutions':'qtr4_contributions','total_qtrly_wages':'qtr4_wages'})
         NYS_industry_2017_4.head()
```

Out[13]:

	year	qtr	industry_code	qtr4_contributions	qtr4_wages	quarter4_emplvl
0	2017	3	10	337029234	146464293659	27685193
1	2017	3	10	0	2225207312	347135
2	2017	3	101	0	143882	15
3	2017	3	1013	0	143882	15
4	2017	3	102	0	2225063430	347120

Combine quarter 1 and quarter 2 into one dataframe. Then add quarter 3, and finally quarter 4.

In [20]:
```
NYS_Industry_2017_12 = pd.merge(NYS_industry_2017_1, NYS_industry_2017_2
, how='inner', left_on='year', right_on='year')
NYS_Industry_2017_12.head()
```

Out[20]:

	year	qtr_x	industry_code_x	qtr1_contributions	qtr1_wages	quarter1_emplvl	qtr_y
0	2017	1	10	1529251900	182980473877	27325310	2
1	2017	1	10	1529251900	182980473877	27325310	2
2	2017	1	10	1529251900	182980473877	27325310	2
3	2017	1	10	1529251900	182980473877	27325310	2
4	2017	1	10	1529251900	182980473877	27325310	2

In [22]:
```
NYS_Industry_2017_12['half1_contributions'] = NYS_Industry_2017_12['qtr1
_contributions'] + NYS_Industry_2017_12['qtr2_contributions']
NYS_Industry_2017_12['half1_wages'] = NYS_Industry_2017_12['qtr1_wages']
 + NYS_Industry_2017_12['qtr2_wages']
NYS_Industry_2017_12['half1_emplvl'] = NYS_Industry_2017_12['quarter1_em
plvl'] + NYS_Industry_2017_12['quarter2_emplvl']
NYS_Industry_2017_12.head()
```

Out[22]:

	year	qtr_x	industry_code_x	qtr1_contributions	qtr1_wages	quarter1_emplvl	qtr_y
0	2017	1	10	1529251900	182980473877	27325310	2
1	2017	1	10	1529251900	182980473877	27325310	2
2	2017	1	10	1529251900	182980473877	27325310	2
3	2017	1	10	1529251900	182980473877	27325310	2
4	2017	1	10	1529251900	182980473877	27325310	2

In [29]:
```
NYS_employment_2017_6mo = NYS_employment_2017_12[['year', 'industry_cod
e','half1_contributions','half1_wages','half11_emplvl']]
NYS_employment_2017_6mo.head()
```

```
---------------------------------------------------------------------
----
NameError                                 Traceback (most recent call l
ast)
<ipython-input-29-ac13516c6aad> in <module>()
----> 1 NYS_employment_2017_6mo = NYS_employment_2017_12[['year', 'indu
stry_code','half1_contributions','half1_wages','half11_emplvl']]
      2 NYS_employment_2017_6mo.head()

NameError: name 'NYS_employment_2017_12' is not defined
```

In [14]:
```
NYS_Industry_2017_34 = pd.merge(NYS_industry_2017_3, NYS_industry_2017_4
, how='inner', left_on='year', right_on='year')
NYS_Industry_2017_34.head()
```

Out[14]:

	year	qtr_x	industry_code_x	qtr3_contributions	qtr3_wages	quarter3_emplvl	qtr_y
0	2017	4	10	267758192	175843272174	28347532	3
1	2017	4	10	267758192	175843272174	28347532	3
2	2017	4	10	267758192	175843272174	28347532	3
3	2017	4	10	267758192	175843272174	28347532	3
4	2017	4	10	267758192	175843272174	28347532	3

In [23]:
```
NYS_Industry_2017_34['half2_contributions'] = NYS_Industry_2017_34['qtr3
_contributions'] + NYS_Industry_2017_34['qtr4_contributions']
NYS_Industry_2017_34['half2_wages'] = NYS_Industry_2017_34['qtr3_wages']
 + NYS_Industry_2017_34['qtr4_wages']
NYS_Industry_2017_34['half2_emplvl'] = NYS_Industry_2017_34['quarter3_em
plvl'] + NYS_Industry_2017_34['quarter4_emplvl']
NYS_Industry_2017_34.head()
```

Out[23]:

	year	qtr_x	industry_code_x	qtr3_contributions	qtr3_wages	quarter3_emplvl	qtr_y
0	2017	4	10	267758192	175843272174	28347532	3
1	2017	4	10	267758192	175843272174	28347532	3
2	2017	4	10	267758192	175843272174	28347532	3
3	2017	4	10	267758192	175843272174	28347532	3
4	2017	4	10	267758192	175843272174	28347532	3

In [15]:
```
NYS_Industry_2017 = pd.merge(NYS_Industry_2017_12, NYS_Industry_2017_34,
 how='inner', left_on='year', right_on='year')
NYS_Industry_2017.head()
```

```
---------------------------------------------------------------------
----
NameError                                 Traceback (most recent call l
ast)
<ipython-input-15-798dea7e4fa1> in <module>()
----> 1 NYS_Industry_2017 = pd.merge(NYS_Industry_2017_12, NYS_Industry
_2017_34, how='inner', left_on='year', right_on='year')
      2 NYS_Industry_2017.head()

NameError: name 'NYS_Industry_2017_12' is not defined
```

Methods for the rest of the data:

1. Combine all informations for the four quarters. Create new columns summarizing contributions, wages, and employment level for the year.
2. Create a new dataframe with only the summarizing columns for the year 2017
3. Sort values for employment levels for the year 2017, take the top 5
4. Sort values for quarterly wages for the year 2017, take the top 5
5. Sort values for contributions for the year 2017, take the top 5

See below example of how I would sort the dataframe. The example below is the first quarter of 2017 sorted 3 ways.

Top 5 industries in New York State according to employment levels:

1.
2.
3.
4.
5.

In [16]:
```
Employment_2017_1 = NYS_industry_2017_1.sort_values(['quarter1_emplvl'],
  ascending=[False])
Employment_2017_1[:5]
```

Out[16]:

	year	qtr	industry_code	qtr1_contributions	qtr1_wages	quarter1_emplvl
0	2017	1	10	1529251900	182980473877	27325310
774	2017	1	10	1526714517	161145407747	23114400
779	2017	1	102	1226934779	147254597140	20646115
784	2017	1	1025	138298283	22922608383	5490079
780	2017	1	1021	288913899	19819940633	4621009

Top 5 industries in New York State according to contributions:

1.
2.
3.
4.
5.

In [17]:
```
Contributions_2017_1 = NYS_employment_2017_1.sort_values(['qtrly_contrib
utions'], ascending=[False])
Contributions_2017_1[:5]
```

Out[17]:

	year	qtr	industry_code	qtrly_contributions	total_qtrly_wages	quarter1_emplvl
0	2017	1	10	1529251900	182980473877	27325310
774	2017	1	10	1526714517	161145407747	23114400
779	2017	1	102	1226934779	147254597140	20646115
783	2017	1	1024	338655018	31526553941	3830616
775	2017	1	101	299779738	13890810607	2468285

Top 5 industries in New York State according to total quarterly wages:

1.
2.
3.
4.
5.

In [18]:
```
Wages_2017_1 = NYS_employment_2017_1.sort_values(['total_qtrly_wages'],
ascending=[False])
Wages_2017_1[:5]
```

Out[18]:

	year	qtr	industry_code	qtrly_contributions	total_qtrly_wages	quarter1_emplvl
0	2017	1	10	1529251900	182980473877	27325310
774	2017	1	10	1526714517	161145407747	23114400
779	2017	1	102	1226934779	147254597140	20646115
782	2017	1	1023	187160049	52970734663	2097920
2215	2017	1	52	140289974	49029398098	1516522