Laboratory work 3

Using Pandas for Data Analysis

Goal: Learning main Pandas features for data analysis.

2. Assignment:

1. Download US Baby Names dataset from the site kaggle.com (https://www.kaggle.com/kaggle/us-baby-names?select=NationalNames.csv)

2. Do exercises according to individual task. For calculating the number of individual task, use the formula

```
N = ord("L") % 5 + 1,
```

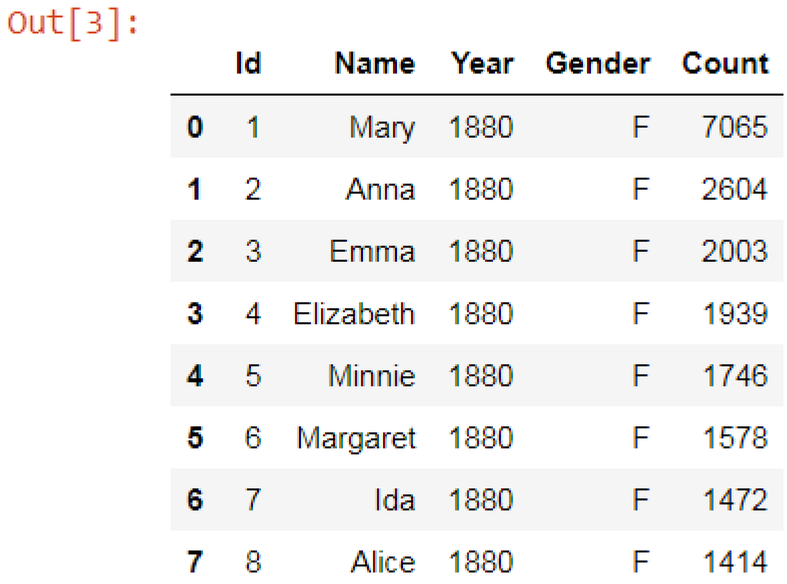
where *N* is number of individual task, *L* is the first letter of your name.

Individual task	Exercises
1	1, 2, 3, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 21, 22, 23, 24, 26
2	3, 4, 5, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20, 22, 23, 24, 27
3	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 18, 19, 20, 21, 23, 25, 27
4	1, 3, 6, 7, 8, 13, 14, 15, 13, 16, 17, 19, 20, 22, 24, 25, 26, 27
5	2, 4, 6, 7, 9, 10, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 26, 27

Exercises

1. Output the first 8 rows of the dataset

Expected output:

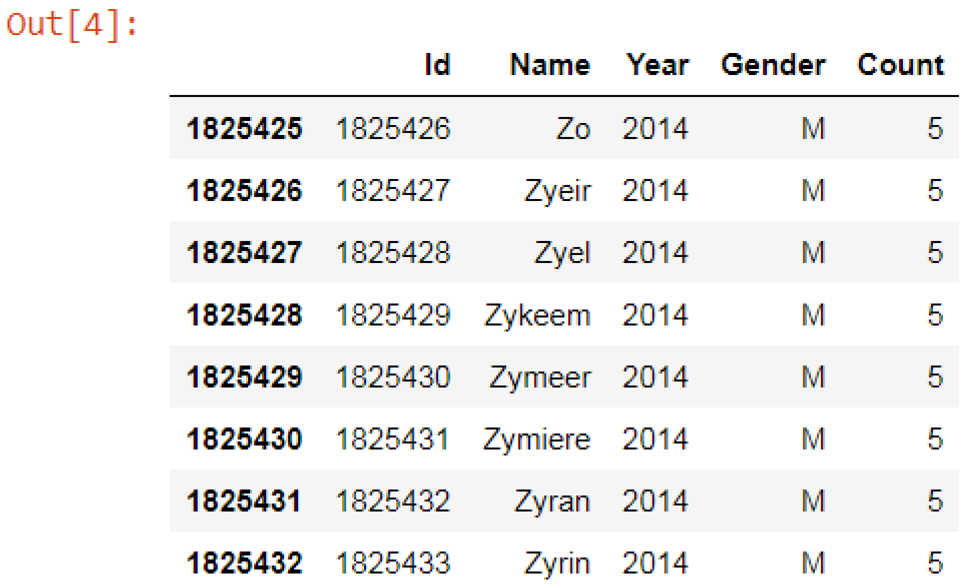


Out[3]:

	Id	Name	Year	Gender	Count
0	1	Mary	1880	F	7065
1	2	Anna	1880	F	2604
2	3	Emma	1880	F	2003
3	4	Elizabeth	1880	F	1939
4	5	Minnie	1880	F	1746
5	6	Margaret	1880	F	1578
6	7	Ida	1880	F	1472
7	8	Alice	1880	F	1414

2. Output the last 8 rows of the dataset

Expected output:

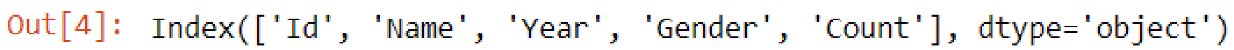


Out[4]:

	Id	Name	Year	Gender	Count
1825425	1825426	Zo	2014	M	5
1825426	1825427	Zyeir	2014	M	5
1825427	1825428	Zyel	2014	M	5
1825428	1825429	Zykeem	2014	M	5
1825429	1825430	Zymeer	2014	M	5
1825430	1825431	Zymiere	2014	M	5
1825431	1825432	Zyran	2014	M	5
1825432	1825433	Zyrin	2014	M	5

3. Get the names of dataset columns

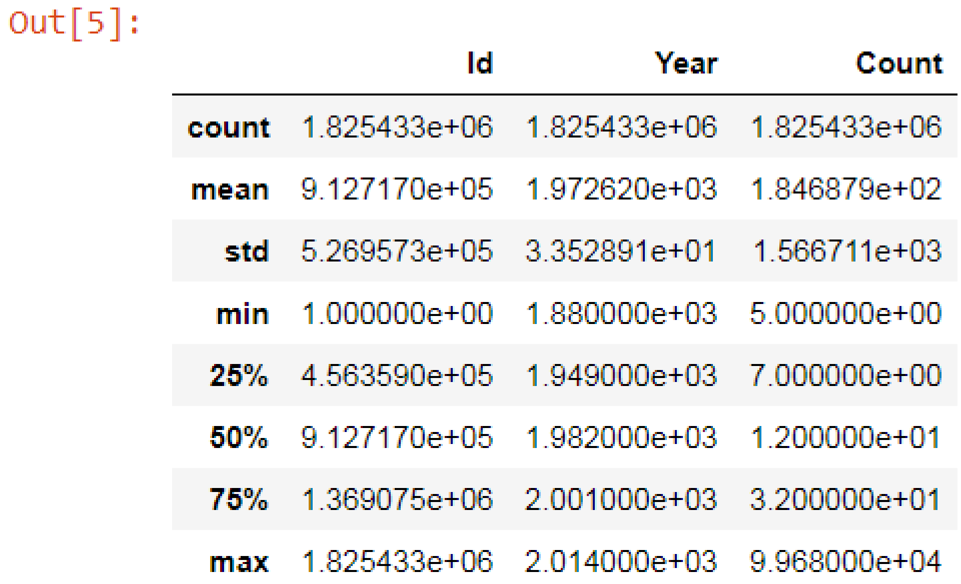
Expected output:



```
Out[4]: Index(['Id', 'Name', 'Year', 'Gender', 'Count'], dtype='object')
```

4. Get general information about data in the dataset

Expected output:

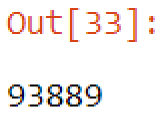


Out[5]:

	Id	Year	Count
count	1.825433e+06	1.825433e+06	1.825433e+06
mean	9.127170e+05	1.972620e+03	1.846879e+02
std	5.269573e+05	3.352891e+01	1.566711e+03
min	1.000000e+00	1.880000e+03	5.000000e+00
25%	4.563590e+05	1.949000e+03	7.000000e+00
50%	9.127170e+05	1.982000e+03	1.200000e+01
75%	1.369075e+06	2.001000e+03	3.200000e+01
max	1.825433e+06	2.014000e+03	9.968000e+04

5. Find the number of unique names in whole dataset

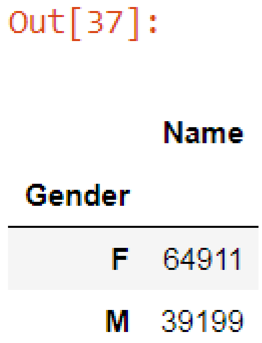
Expected output:



```
Out[33]:

93889
```

6. Calculate the number of unique female and male names in whole dataset

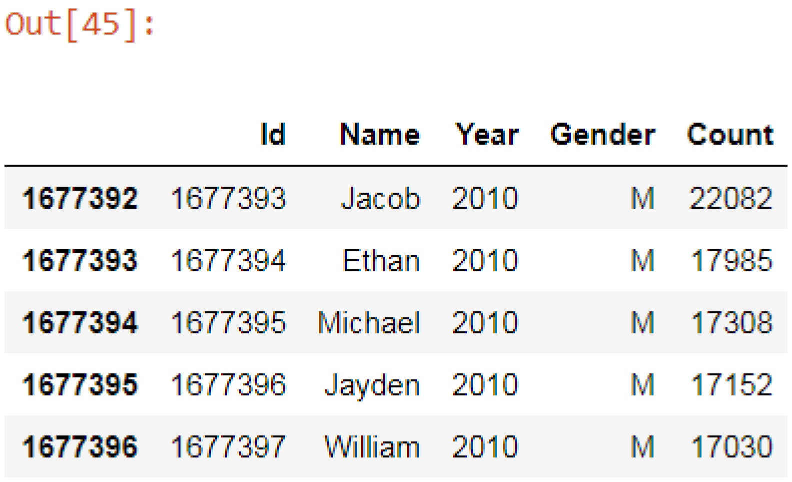


Out[37]:

	Name
Gender	
F	64911
M	39199

7. Find 5 the most popular male names in 2010

Expected output:

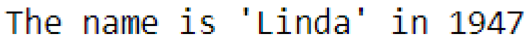


Out[45]:

	Id	Name	Year	Gender	Count
1677392	1677393	Jacob	2010	M	22082
1677393	1677394	Ethan	2010	M	17985
1677394	1677395	Michael	2010	M	17308
1677395	1677396	Jayden	2010	M	17152
1677396	1677397	William	2010	M	17030

8. Find the most popular name based on the results of one year (the name for which `Count` is maximum)

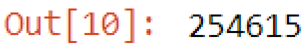
Expected output:



```
The name is 'Linda' in 1947
```

9. Count the number of records with `Count` = minimum.

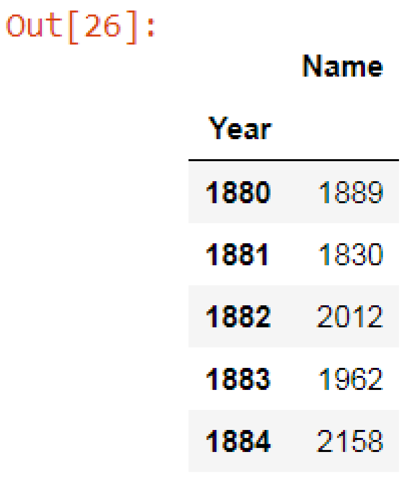
Expected output:



```
Out[10]: 254615
```

10. Count the number of unique names in each year

Expected output:

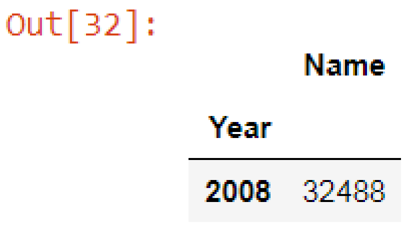


Out[26]:

	Name
Year	
1880	1889
1881	1830
1882	2012
1883	1962
1884	2158

11. Find the year with the most number of unique names.

Expected output:

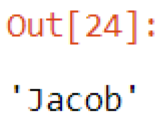


```
Out[32]:
```

	Name
Year	
2008	32488

12. Find most popular name of the year with the most number of unique names (that is in 2008)

Expected output:

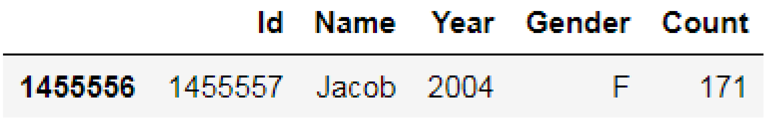


```
Out[24]:

'Jacob'
```

13. Find the year when the name “Jacob” was the most popular as a female name

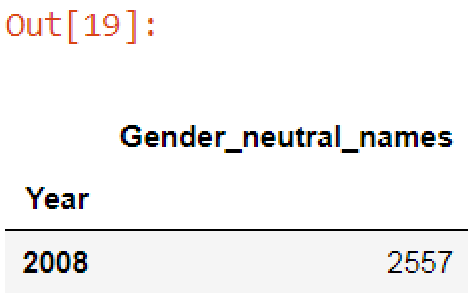
Expected output:



	Id	Name	Year	Gender	Count
1455556	1455557	Jacob	2004	F	171

14. Find year, with the most number of gender neutral names (the same male and female names)

Expected output:

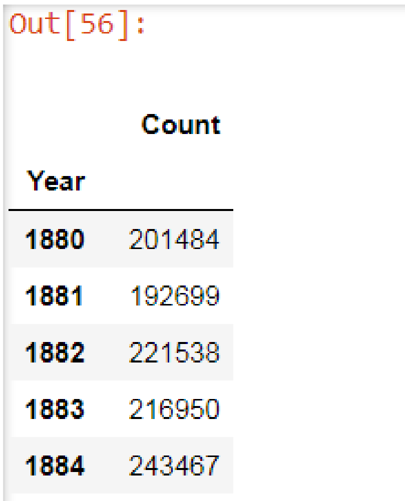


```
Out[19]:
```

	Gender_neutral_names
Year	
2008	2557

15. Find total births per year

Expected output of the first 5 rows:

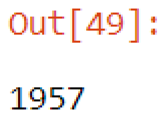


Out[56]:

	Count
Year	
1880	201484
1881	192699
1882	221538
1883	216950
1884	243467

16. Find the year when the greatest number of children was born

Expected output:

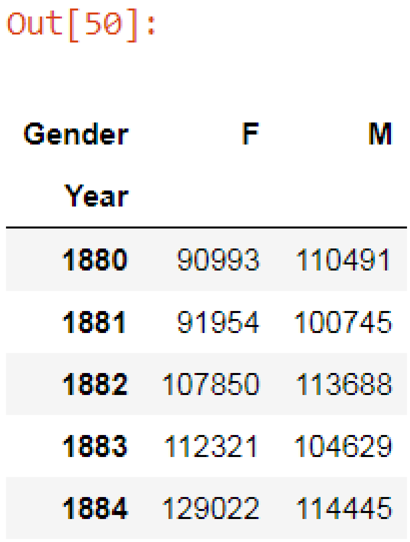


```
Out[49]:

1957
```

17. Find the number of girls and boys that were born in each year

Expected output of the first 5 rows:

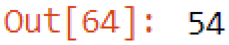


Out[50]:

Gender	**F**	**M**
Year		
1880	90993	110491
1881	91954	100745
1882	107850	113688
1883	112321	104629
1884	129022	114445

18. Count the number of years when more girls were born than boys

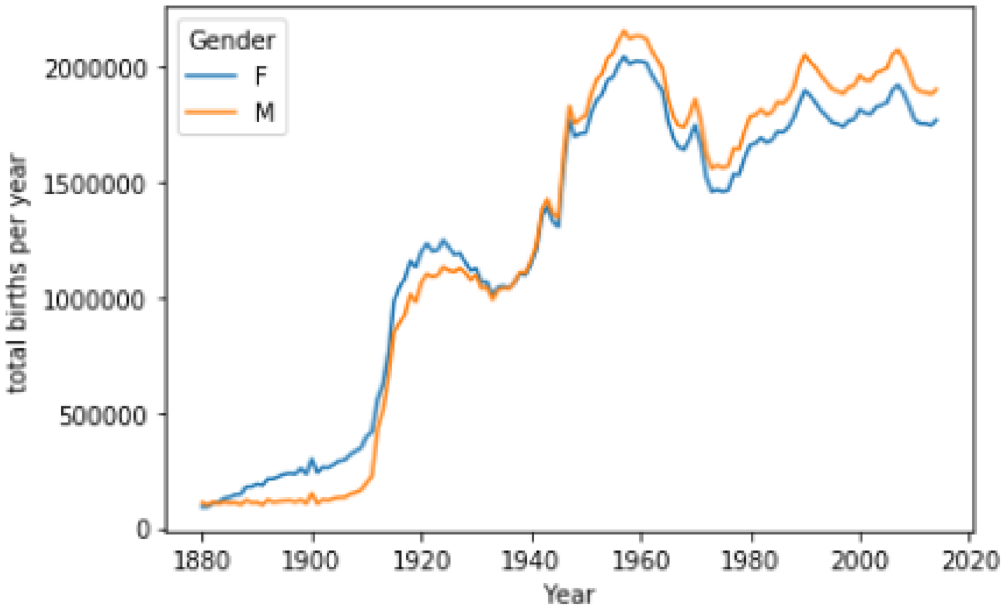
Expected output:



```
Out[64]: 54
```

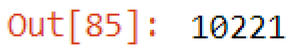
19. Draw the plot of total births per year of boys and girls

Expected output:



20. Count number of gender neutral names (same for girls and boys)

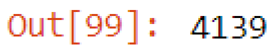
Expected output:



```
Out[85]:  10221
```

21. Count how much times boys were named as Barbara

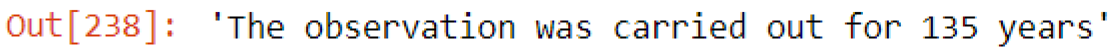
Expected output:



```
Out[99]:  4139
```

22. Calculate how many years the observation was carried out

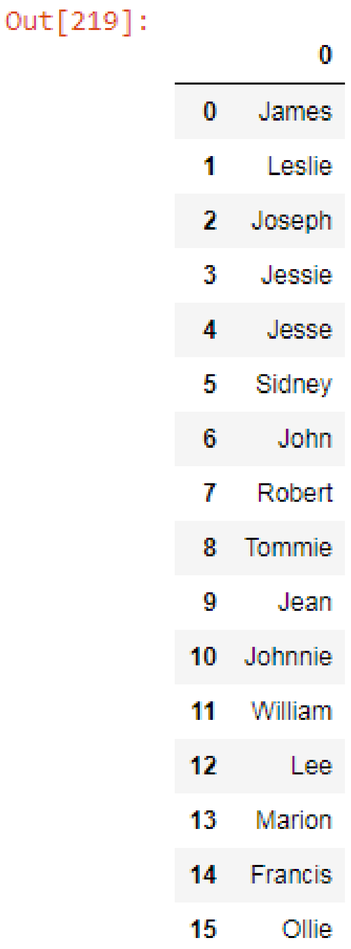
Expected output:



```
Out[238]:  'The observation was carried out for 135 years'
```

23. Find the most popular gender neutral names (those present each year)

Expected output:

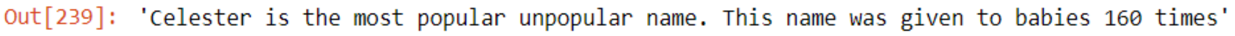


Out[219]:

	0
0	James
1	Leslie
2	Joseph
3	Jessie
4	Jesse
5	Sidney
6	John
7	Robert
8	Tommie
9	Jean
10	Johnnie
11	William
12	Lee
13	Marion
14	Francis
15	Ollie

24. Find the most popular unpopular names (unpopular name that babies have been called the most times)

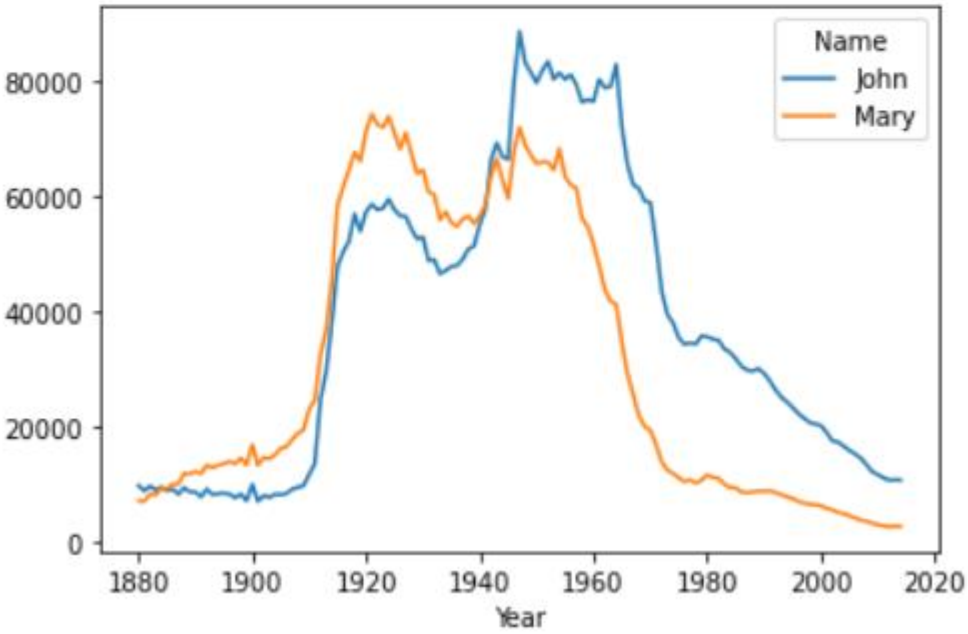
Expected output:



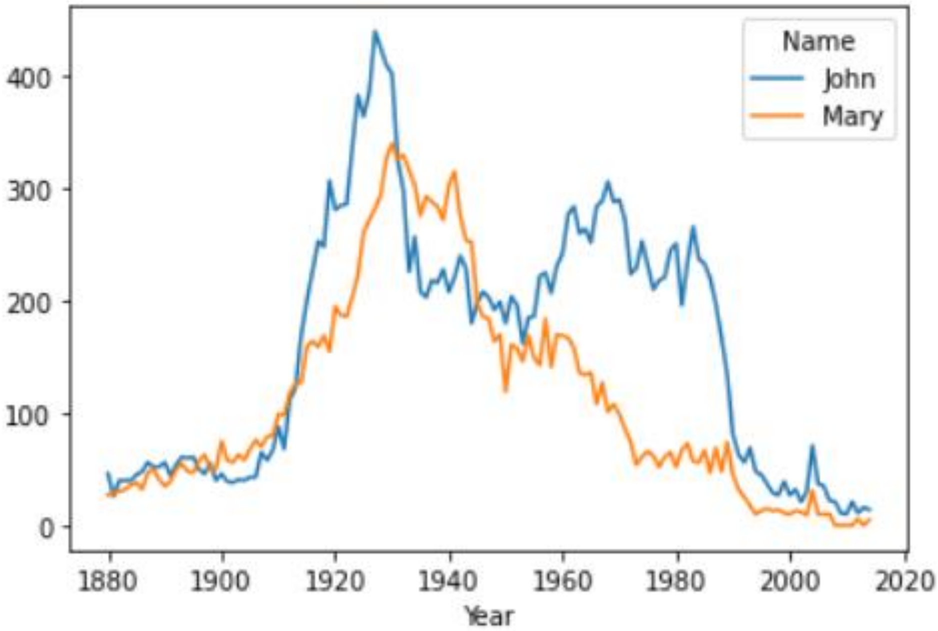
```
Out[239]: 'Celester is the most popular unpopular name. This name was given to babies 160 times'
```

25. Plot graphs of the distribution of the number of names “John” and “Mary” by years, regardless of gender.

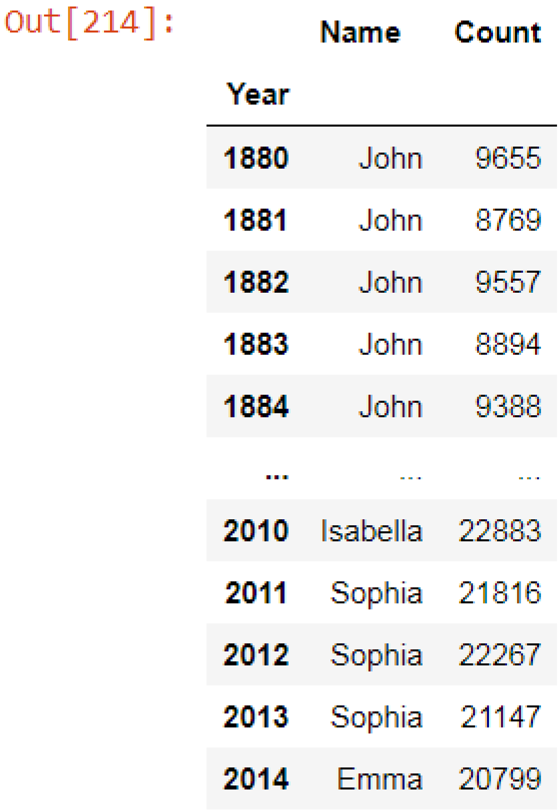
Expected output:



26. Plot graphs of the distribution of the number of female names “John” and male names “Mary” by years



27. Find the most popular names each year



Out[214]:

	Name	Count
Year		
1880	John	9655
1881	John	8769
1882	John	9557
1883	John	8894
1884	John	9388
...	...	...
2010	Isabella	22883
2011	Sophia	21816
2012	Sophia	22267
2013	Sophia	21147
2014	Emma	20799

3. The content of the report

1. Cover page of the report.
2. Topic and goal of the lab.
3. Progress of the work.
4. Link to the created Jupyter Notebook on GitHub, rendered by nbviewer.
5. Conclusions.