

## Read the dataset from the below link

[https://raw.githubusercontent.com/guipsamora/pandas\\_exercises/master/06\\_Scatter\\_Plot/US\\_Baby\\_Names/US\\_Baby\\_Names\\_right.csv](https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/06_Scatter_Plot/US_Baby_Names/US_Baby_Names_right.csv)

```
In [1]: import pandas as pd
%matplotlib inline
df=pd.read_csv('https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/06_Scatter_Plot/US_Baby_Names/US_Baby_Names_right.csv')
df.head()
```

Out[1]:

	Unnamed: 0	Id	Name	Year	Gender	State	Count
0	11349	11350	Emma	2004	F	AK	62
1	11350	11351	Madison	2004	F	AK	48
2	11351	11352	Hannah	2004	F	AK	46
3	11352	11353	Grace	2004	F	AK	44
4	11353	11354	Emily	2004	F	AK	41

## 1. Delete unnamed columns

```
In [2]: df.drop(df[df.columns[df.columns.str.contains('Unnamed: ')]],axis=1,inplace=True)
df.head()
```

Out[2]:

	Id	Name	Year	Gender	State	Count
0	11350	Emma	2004	F	AK	62
1	11351	Madison	2004	F	AK	48
2	11352	Hannah	2004	F	AK	46
3	11353	Grace	2004	F	AK	44
4	11354	Emily	2004	F	AK	41

## 2. Show the distribution of male and female

```
In [3]: #Bar with count of Male and Female
#df.plot.bar("Gender")
df.groupby("Gender").describe()
```

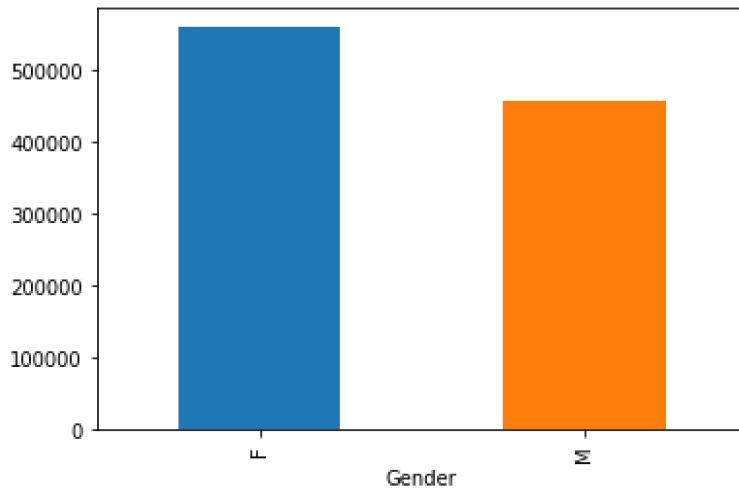
Out[3]:

	Count								Id		
	count	mean	std	min	25%	50%	75%	max	count	mean	..
Gender											
F	558846.0	29.310925	75.962992	5.0	6.0	10.0	23.0	3634.0	558846.0	2.793178e+06	.
M	457549.0	41.615650	118.074308	5.0	7.0	12.0	29.0	4167.0	457549.0	2.877176e+06	.

2 rows × 24 columns

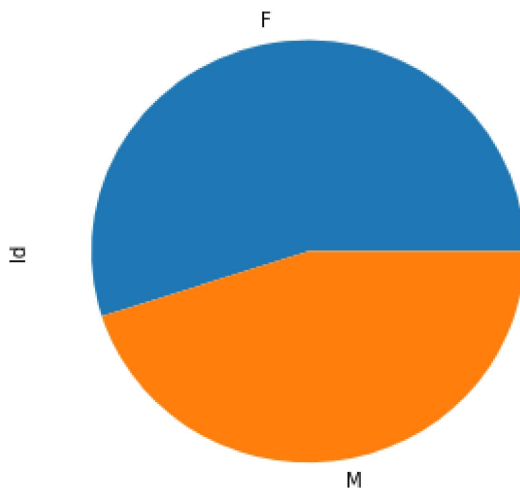
```
In [4]: import matplotlib.pyplot as plt
df.groupby("Gender").count().Id.plot(kind='bar')
```

```
Out[4]: <matplotlib.axes._subplots.AxesSubplot at 0x22c0005bba8>
```



```
In [5]: df.groupby("Gender").count().Id.plot.pie(figsize=(5,5))
```

```
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x22c001ad0f0>
```



### 3. Show the top 5 most preferred names

```
In [6]: df.Name.value_counts(sort=True).head(5)
```

```
Out[6]: Riley      1112
Avery      1080
Jordan     1073
Peyton     1064
Hayden     1049
Name: Name, dtype: int64
```

### 4. What is the median name occurrence in the dataset

```
In [7]: df.Name.value_counts()[df.Name.value_counts() == (df.Name.value_counts().median(
```

```
Out[7]: Daphnie      8
        Lucila      8
        Janeli      8
        Nazier      8
        Adira       8
        Thierno     8
        Fidencio    8
        Ameir       8
        Siyona      8
        Emmaly      8
        Jakhari     8
        Candelario  8
        Miki        8
        Nazar       8
        Jalani      8
        Judas       8
        Joshue      8
        Maurilio    8
        Taisha      8
        Yared       8
        Willian     8
        Noora       8
        Adylene     8
        Ediel       8
        Pasquale    8
        Alliana     8
        Ysabelle    8
        Madix       8
        Yoni        8
        Amyri       8
        ..
        Levin      8
        Kayly       8
        Amaryllis   8
        Jediah      8
        Irais       8
        Armonie     8
        Jatziri     8
        Ridwan      8
        Abubakar    8
        Braidyn     8
        Nassir      8
        Evanna      8
        Ivyanna     8
        Ivett       8
        Aleana      8
        Audrinna    8
        Odelia      8
        Kathlyn     8
        Kimia       8
        Lucienne    8
        Eliya       8
        Baily       8
        Marquell    8
        Kenli       8
        Katiana     8
        Itzell      8
        Gaelle      8
        Aerial      8
        Sabino      8
```

Railynn 8

Name: Name, Length: 360, dtype: int64

## 5. Distribution of male and female born count by states

```
In [8]: df.groupby(["State", "Gender"]).Id.count()
```

```
Out[8]: State Gender
AK      F      2404
        M      2587
AL      F      9878
        M      8419
AR      F      7171
        M      6475
AZ      F     14518
        M     10820
CA      F     45144
        M     31637
CO      F     11424
        M      9183
CT      F      6575
        M      5733
DC      F      3053
        M      3000
DE      F      2549
        M      2440
FL      F     25781
        M     20070
GA      F     19385
        M     15454
HI      F      3255
        M      3546
IA      F      7131
        M      6307
ID      F      4918
        M      4833
IL      F     21268
        M     16828
        ...
OK      F      9519
        M      8138
OR      F      8604
        M      7333
PA      F     17480
        M     14171
RI      F      2558
        M      2468
SC      F      9465
        M      8195
SD      F      2838
        M      2908
TN      F     13063
        M     10588
TX      F     39760
        M     27791
UT      F      9515
        M      8233
VA      F     14759
        M     11997
VT      F      1398
        M      1618
WA      F     13329
        M     11049
WI      F     10549
        M      8940
WV      F      4305
        M      3733
```

```
WY      F      1456
        M      1904
Name: Id, Length: 102, dtype: int64
```

In [ ]: