

In [1]:

```
import pandas as pd
import numpy as np
import os

data = pd.read_csv("https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/06_Stats/US_Baby_Names/US_Baby_Names_right.csv")

data.head(5)
```

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

In [2]:

```
data.drop(['Unnamed: 0'], axis=1, inplace=True)

data.head(5)
```

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

In [3]:

```
data_names = data.groupby("Name").sum()  
data_names.sort_values("Count", ascending = 0).head(5)
```

Out[3]:

	Id	Year	Count
Name			
Jacob	1665681356	1141099	242874
Emma	1629482250	1137085	214852
Michael	1687521295	1161152	214405
Ethan	1660808475	1139091	209277
Isabella	1630131786	1137090	204798

In [4]:

```
data_names[data_names.Count == data_names.Count.median()]
```

Out[4]:

	Id	Year	Count
Name			
Aishani	7810526	14078	49
Alara	18841027	16079	49
Alysse	22629405	16057	49
Ameir	21780411	16086	49
Anely	4349541	16071	49
Antonina	27672250	18081	49
Aveline	7982905	12065	49
Aziah	29825407	16073	49
Baily	27406186	16064	49
Caleah	20967785	18106	49
Carlota	6971174	14077	49
Cristine	11299091	14042	49
Dahlila	8183033	14063	49
Darvin	13433473	16078	49

Deante	24229110	18064	49
Deserae	31068418	18061	49
Devean	7100057	8019	49
Elizah	13583872	16063	49
Emmaly	17488711	16075	49
Emmanuela	26771519	18074	49
Envy	13006001	16070	49
Esli	17424928	16059	49
Fay	17137619	16072	49
Gurshaan	4884930	14070	49
Hareem	16374682	14090	49
Iven	4861915	14062	49
Jaice	31448007	16098	49
Jaiyana	23805956	14068	49
Jamiracle	16247591	18091	49
Jelissa	24630716	16070	49
...
Kyndle	33887372	16082	49
Kynsley	26646473	14084	49
Leylanie	4347491	16070	49
Maisha	23441764	14047	49
Malillany	17027066	14087	49
Mariann	22330712	16060	49
Marquell	18281388	16053	49
Maurilio	22999235	16058	49
Mckynzie	30075282	14068	49
Mehdi	23762227	16070	49
Nabeel	21778878	18070	49
Nalleli	17420939	16058	49
Nassir	18112868	16058	49
Nazier	31970176	16061	49
Nishant	12096448	16050	49
Rebecka	17535450	16061	49

Reghan	31154500	18073	49
Ridwan	25287426	16082	49
Riot	26477266	16104	49
Rubin	18701425	16055	49
Ryatt	25392242	16103	49
Sameera	27250241	18083	49
Sanjuanita	24554019	10035	49
Shalyn	29474001	18061	49
Skylie	27316494	16086	49
Sriram	9208203	14054	49
Trinton	32822302	16069	49
Vita	10241632	14075	49
Yoni	6439154	16060	49
Zuleima	3521373	14050	49

66 rows × 3 columns

In [5]:

```
data['State'].value_counts()
```

Out[5]:

CA	76781
TX	67551
NY	50743
FL	45851
IL	38096
GA	34839
OH	32461
PA	31651
NC	30887
MI	29281
NJ	27315
VA	26756
AZ	25338
WA	24378
IN	23669
TN	23651
MO	21865
MD	20759
CO	20607
LA	20186
MN	19681
WI	19489
MA	19189
AL	18297
UT	17748
SC	17660
OK	17657
KY	16084
OR	15937
KS	14501
MS	14097
AR	13646
IA	13438
NV	13116
CT	12308
NM	10687
NE	10399
ID	9751
WV	8038
HI	6801
DC	6053
ME	5753
SD	5746
MT	5676
NH	5616
RI	5026
AK	4991
DE	4989
ND	4980
WY	3360
VT	3016

Name: State, dtype: int64

In []: