In [10]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
df=pd.read_csv('https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/06_Stats/US_Baby_Names/US_Baby_Names_right.csv')
df.head()
```

Out[10]:

	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]:
```
#1. Delete unnamed columns
df.drop(df.columns[df.columns.str.contains('Unnamed',case = False)], 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

In [11]:
```
#2. Show the distribution of male and female

np.round((df['Gender'].value_counts())/len(df)*100,2)
```

Out[11]:
```
F    54.98
M    45.02
Name: Gender, dtype: float64
```

In [16]:
```
#3. Show the top 5 most preferred names

#Grouping the values on basis of name and count and taking a sum followed by s
orting in descending order to get the most
#common five names in the data set.

df.groupby('Name')['Count'].sum().sort_values(ascending=False).head(5)
```

Out[16]:
```
Name
Jacob       242874
Emma        214852
Michael     214405
Ethan       209277
Isabella    204798
Name: Count, dtype: int64
```

In [18]:
```
# 4. What is the median name occurence in the dataset

#Finding the median value using Id
df.median()['Id']
```

Out[18]:
```
2811921.0
```

In [37]:
```
print('The median name occuring in the data set is : ')
#Populating the respecitve name for the median Id
df[df['Id'] == df.median()['Id']]['Name']
```

```
The median name occuring in the data set is :
```

Out[37]:
```
508197    Kasey
Name: Name, dtype: object
```

In [31]:
```
# 5. Distribution of male and female born count by states
#Grouping the data by state and gender
df.groupby(['State','Gender'])['Count'].sum()
```

Out[31]:

State	Gender	
AK	F	26250
	M	37399
AL	F	215308
	M	260114
AR	F	129712
	M	162947
AZ	F	368567
	M	439691
CA	F	2414063
	M	2670584
CO	F	260805
	M	313425
CT	F	141350
	M	171397
DC	F	35276
	M	47228
DE	F	31312
	M	41748
FL	F	915422
	M	1060957
GA	F	549637
	M	635531
HI	F	37279
	M	53127
IA	F	144764
	M	174009
ID	F	72808
	M	94320
IL	F	695312
	M	791679
		...
OK	F	184967
	M	228613
OR	F	172111
	M	209445
PA	F	593382
	M	682709
RI	F	35560
	M	47939
SC	F	197917
	M	237442
SD	F	34104
	M	45443
TN	F	336487
	M	398615
TX	F	1786281
	M	2005394
UT	F	202892
	M	245324
VA	F	405503
	M	466873
VT	F	15079
	M	21353
WA	F	334944
	M	395377
WI	F	264921

```
       M          311758
WV     F           73800
       M           93557
WY     F           14107
       M           21912
Name: Count, Length: 102, dtype: int64
```

In [39]:
```
graph = df.groupby(['State','Gender'])['Count'].sum()
print('-------- Gender wise distribution of population across states ---------
')
graph.unstack().plot(kind='bar',width=0.8,stacked=True,  color=['Orange','Blu
e'], grid=False,figsize=(15,5))
```

```
-------- Gender wise distribution of population across states ---------
```

Out[39]: <matplotlib.axes._subplots.AxesSubplot at 0xf0c6f70>