Ques 09 : When should I use a "GROUP BY" pandas?

In [10]:
```
import pandas as pd
drinks_ds = pd.read_csv('http://bit.ly/drinksbycountry')
drinks_ds.head(4)
```

Out[10]:

	country	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol	continent
0	Afghanistan	0	0	0	0.0	Asia
1	Albania	89	132	54	4.9	Europe
2	Algeria	25	0	14	0.7	Africa
3	Andorra	245	138	312	12.4	Europe

In [12]:
```
# What is the average beer_servings across all continents?
drinks_ds.beer_servings.mean()
```

Out[12]:
```
106.16062176165804
```

In [13]:
```
# What is the average beer_servings by continents? How beer_servings varied from continent to continent ?
drinks_ds.groupby('continent').beer_servings.mean()
```

Out[13]:
```
continent
Africa            61.471698
Asia              37.045455
Europe           193.777778
North America    145.434783
Oceania           89.687500
South America    175.083333
Name: beer_servings, dtype: float64
```

In [5]:
```
# Max beer_servings by continents
drinks_ds.groupby('continent').beer_servings.max()
```

Out[5]:
```
continent
Africa           376
Asia             247
Europe           361
North America    285
Oceania          306
South America    333
Name: beer_servings, dtype: int64
```

Specifying multiple aggregation functions at once

In [6]:
```
drinks_ds.groupby('continent').beer_servings.agg(['count','max','min','mean'])
```

Out[6]:

continent	count	max	min	mean
Africa	53	376	0	61.471698
Asia	44	247	0	37.045455
Europe	45	361	0	193.777778
North America	23	285	1	145.434783
Oceania	16	306	0	89.687500
South America	12	333	93	175.083333

Calculating mean/max/min for all numeric columns

In [7]:
```
drinks_ds.groupby('continent').mean()
```

Out[7]:

continent	beer_servings	spirit_servings	wine_servings	total_litres_of_pure_alcohol
Africa	61.471698	16.339623	16.264151	3.007547
Asia	37.045455	60.840909	9.068182	2.170455
Europe	193.777778	132.555556	142.222222	8.617778
North America	145.434783	165.739130	24.521739	5.995652
Oceania	89.687500	58.437500	35.625000	3.381250
South America	175.083333	114.750000	62.416667	6.308333

In [8]:
```
drinks_ds.groupby('continent').agg(['count','mean','max','min'])
```

Out[8]:

	beer_servings				spirit_servings				wine_servings				total_litres_of_pure_alcohol			
	count	mean	max	min	count	mean	max	min	count	mean	max	min	count	mean	max	min
continent																
Africa	53	61.471698	376	0	53	16.339623	152	0	53	16.264151	233	0	53	3.007547	9.1	0.0
Asia	44	37.045455	247	0	44	60.840909	326	0	44	9.068182	123	0	44	2.170455	11.5	0.0
Europe	45	193.777778	361	0	45	132.555556	373	0	45	142.222222	370	0	45	8.617778	14.4	0.0
North America	23	145.434783	285	1	23	165.739130	438	68	23	24.521739	100	1	23	5.995652	11.9	2.2
Oceania	16	89.687500	306	0	16	58.437500	254	0	16	35.625000	212	0	16	3.381250	10.4	0.0
South America	12	175.083333	333	93	12	114.750000	302	25	12	62.416667	221	1	12	6.308333	8.3	3.8

Visually showing the result

In [9]:
```
%matplotlib inline

drinks_ds.groupby('continent').mean().plot(kind='bar')
```

Out[9]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x55145c3860>
```