In [1]:
```
import pandas as pd
# load the tsv file
dataframe = pd.read_csv('data/gapminder.tsv',sep='\t')

#print first 5 rows
print (dataframe.head())
```

```
       country continent  year  lifeExp       pop   gdpPercap
0  Afghanistan      Asia  1952   28.801   8425333  779.445314
1  Afghanistan      Asia  1957   30.332   9240934  820.853030
2  Afghanistan      Asia  1962   31.997  10267083  853.100710
3  Afghanistan      Asia  1967   34.020  11537966  836.197138
4  Afghanistan      Asia  1972   36.088  13079460  739.981106
```

In [2]:
```
# print the number of rows and coloumns in the tsv file
print (dataframe.shape)
```

```
(1704, 6)
```

In [3]:
```
# print the name of the coloumns
print (dataframe.columns) # columns is an attribute not a method
```

```
Index(['country', 'continent', 'year', 'lifeExp', 'pop', 'gdpPercap'], dtype='ob
ject')
```

In [4]:
```
# print the type of each column
print (dataframe.dtypes)
```

```
country       object
continent     object
year           int64
lifeExp      float64
pop            int64
gdpPercap    float64
dtype: object
```

In [5]:
```
# most common datatype using pandas
# object --> string
# int64 --> int
#float64 --> float
#datetime64 --> datetime
```

In [6]:
```
# get the of a single column
dataframe_country = dataframe['country']
print (dataframe_country.head()) # print first 5 rows
```

```
0    Afghanistan
1    Afghanistan
2    Afghanistan
3    Afghanistan
4    Afghanistan
Name: country, dtype: object
```

In [7]:
```
# shows the last 5 data from the dataframe
print (dataframe.tail())
```

```
       country continent  year  lifeExp       pop   gdpPercap
1699  Zimbabwe    Africa  1987   62.351   9216418  706.157306
1700  Zimbabwe    Africa  1992   60.377  10704340  693.420786
1701  Zimbabwe    Africa  1997   46.809  11404948  792.449960
1702  Zimbabwe    Africa  2002   39.989  11926563  672.038623
1703  Zimbabwe    Africa  2007   43.487  12311143  469.709298
```

In [8]:
```
# print the last 5 country name from
print (dataframe_country.tail())
```

```
1699    Zimbabwe
1700    Zimbabwe
1701    Zimbabwe
1702    Zimbabwe
1703    Zimbabwe
Name: country, dtype: object
```

In [10]:
```
# store x number of columns value in a variable
store_data = dataframe[['country','continent','year']]

# print first 5 rows from the store_data vairbale
print (store_data.head())
```

```
       country continent  year
0  Afghanistan      Asia  1952
1  Afghanistan      Asia  1957
2  Afghanistan      Asia  1962
3  Afghanistan      Asia  1967
4  Afghanistan      Asia  1972
```

In [4]:
```
# get the first row
import pandas as pd
dataframe = pd.read_csv('data/gapminder.tsv',sep='\t')
store_data = dataframe[['country','continent','year']]
print (store_data.iloc[0])
```

```
country      Afghanistan
continent           Asia
year                1952
Name: 0, dtype: object
```

In [5]:
```
# get the last row index and then display the data
number_of_rows = dataframe.shape[0]
print(number_of_rows)
last_row_index = number_of_rows - 1

#print data in the last row
print (dataframe.loc[last_row_index])
```

```
1704
country      Zimbabwe
continent      Africa
year             2007
lifeExp        43.487
pop          12311143
gdpPercap     469.709
Name: 1703, dtype: object
```

In [6]:
```
# subsetting columns
subset = dataframe.loc[:,['year','pop']]
print (subset.head())
```

```
   year       pop
0  1952   8425333
1  1957   9240934
2  1962  10267083
3  1967  11537966
4  1972  13079460
```

In [7]:
```
#display the last column
subset = dataframe.iloc[:,[2,4,-1]]
print(subset.head())
```

```
   year       pop   gdpPercap
0  1952   8425333  779.445314
1  1957   9240934  820.853030
2  1962  10267083  853.100710
3  1967  11537966  836.197138
4  1972  13079460  739.981106
```

In [8]:
```
# Coulmn slicing
subset = dataframe.iloc[:,:3]
print(subset.head())
```

```
       country continent  year
0  Afghanistan      Asia  1952
1  Afghanistan      Asia  1957
2  Afghanistan      Asia  1962
3  Afghanistan      Asia  1967
4  Afghanistan      Asia  1972
```

In [9]:
```
# selecting specific rows with specific column
print(dataframe.loc[30,'country'])
```

```
Algeria
```

In [10]:
```
print(dataframe.loc[30,'year'])
```

```
1982
```

In [11]:
```
# The same above operation using iloc
print(dataframe.iloc[42,0])
```

```
Angola
```

In [12]:
```
print(dataframe.iloc[42,1])
```

```
Africa
```

In [13]:
```
# Grouped or aggregate calculation
print(dataframe.groupby('year')['lifeExp'].mean())
```

```
year
1952    49.057620
1957    51.507401
1962    53.609249
1967    55.678290
1972    57.647386
1977    59.570157
1982    61.533197
1987    63.212613
1992    64.160338
1997    65.014676
2002    65.694923
2007    67.007423
Name: lifeExp, dtype: float64
```

In [14]:
```
print(dataframe.groupby(['year','continent'])[['lifeExp','gdpPercap']].mean())
```

year	continent	lifeExp	gdpPercap
1952	Africa	39.135500	1252.572466
	Americas	53.279840	4079.062552
	Asia	46.314394	5195.484004
	Europe	64.408500	5661.057435
	Oceania	69.255000	10298.085650
1957	Africa	41.266346	1385.236062
	Americas	55.960280	4616.043733
	Asia	49.318544	5787.732940
	Europe	66.703067	6963.012816
	Oceania	70.295000	11598.522455
1962	Africa	43.319442	1598.078825
	Americas	58.398760	4901.541870
	Asia	51.563223	5729.369625
	Europe	68.539233	8365.486814
	Oceania	71.085000	12696.452430
1967	Africa	45.334538	2050.363801
	Americas	60.410920	5668.253496
	Asia	54.663640	5971.173374
	Europe	69.737600	10143.823757
	Oceania	71.310000	14495.021790
1972	Africa	47.450942	2339.615674
	Americas	62.394920	6491.334139
	Asia	57.319269	8187.468699
	Europe	70.775033	12479.575246
	Oceania	71.910000	16417.333380
1977	Africa	49.580423	2585.938508
	Americas	64.391560	7352.007126
	Asia	59.610556	7791.314020
	Europe	71.937767	14283.979110
	Oceania	72.855000	17283.957605
1982	Africa	51.592865	2481.592960
	Americas	66.228840	7506.737088
	Asia	62.617939	7434.135157
	Europe	72.806400	15617.896551
	Oceania	74.290000	18554.709840
1987	Africa	53.344788	2282.668991
	Americas	68.090720	7793.400261
	Asia	64.851182	7608.226508
	Europe	73.642167	17214.310727
	Oceania	75.320000	20448.040160
1992	Africa	53.629577	2281.810333
	Americas	69.568360	8044.934406
	Asia	66.537212	8639.690248
	Europe	74.440100	17061.568084
	Oceania	76.945000	20894.045885
1997	Africa	53.598269	2378.759555
	Americas	71.150480	8889.300863
	Asia	68.020515	9834.093295
	Europe	75.505167	19076.781802
	Oceania	78.190000	24024.175170
2002	Africa	53.325231	2599.385159
	Americas	72.422040	9287.677107
	Asia	69.233879	10174.090397
	Europe	76.700600	21711.732422
	Oceania	79.740000	26938.778040
2007	Africa	54.806038	3089.032605
	Americas	73.608120	11003.031625
	Asia	70.728485	12473.026870
	Europe	77.648600	25054.481636
	Oceania	80.719500	29810.188275

In [15]:
```
# Grouped Frequency Counts
print(dataframe.groupby('continent')['country'].nunique())
```

```
continent
Africa      52
Americas    25
Asia        33
Europe      30
Oceania      2
Name: country, dtype: int64
```

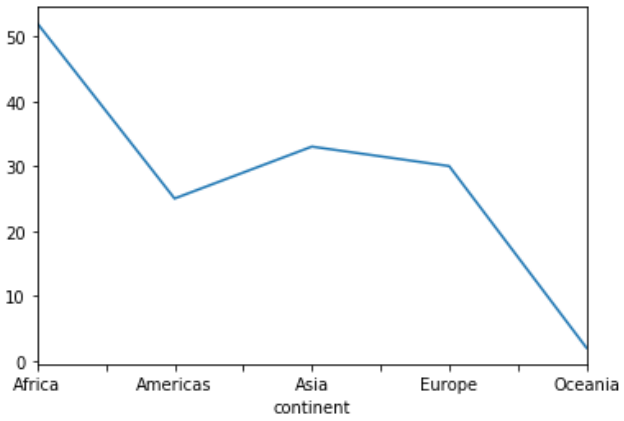
In [16]:
```
# Basic Plot
country_plot = dataframe.groupby('continent')['country'].nunique()
country_plot.plot()
```

Out[16]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7faa9419a160>
```

In [17]:
```
country_plot.plot().show()
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
<ipython-input-17-1e20e7dde11e> in <module>
----> 1 country_plot.plot().show()

AttributeError: 'AxesSubplot' object has no attribute 'show'
```



In [ ]:
```

```