# 시리즈 데이터 만들기

- pd.Series()

```
import pandas as pd

series_data1 = pd.Series([1, 2, 3])
series_data2 = pd.Series(('a', 'b', 'c'))
series_data3 = pd.Series({'a': 1, 'b': 2, 'c': 3})
```

```
series_data1
```

```
0    1
1    2
2    3
dtype: int64
```

```
series_data2
```

```
0    a
1    b
2    c
dtype: object
```

```
series_data3
```

```
a    1
b    2
c    3
dtype: int64
```

# 데이터프레임 만들기

- pd.DataFrame(리스트, columns=[열이름])
- pd.DataFrame({키:값})
  ↭ 키가 열이름이 된다.

```
li = [5, 10, 15]
df = pd.DataFrame(li, columns=['number'])
df
```

| | number |
|---|---|
| 0 | 5 |
| 1 | 10 |
| 2 | 15 |

```
li = [[7, 8, 9], [10, 10, 9]]
df = pd.DataFrame(li, columns=['score1', 'score2', 'score3'])
df
```

| | score1 | score2 | score3 |
|---|---|---|---|
| 0 | 7 | 8 | 9 |
| 1 | 10 | 10 | 9 |

```
df = pd.DataFrame([
    {'name': 'apple', 'price': 30, 'quantity': 120},
    {'name': 'pitch', 'price': 40, 'quantity': 100}])
df
```

| | name | price | quantity |
|---|---|---|---|
| 0 | apple | 30 | 120 |
| 1 | pitch | 40 | 100 |

# 데이터 파일 읽기

- pd.read_csv('파일명.csv')

```
import pandas as pd

df=pd.read_csv('sample_data/california_housing_test.csv')

df.head()
```

| | longitude | latitude | housing_median_age | total_rooms | total_bedrooms | populatio |
|---|---|---|---|---|---|---|
| 0 | -122.05 | 37.37 | 27.0 | 3885.0 | 661.0 | 1537. |
| 1 | -118.30 | 34.26 | 43.0 | 1510.0 | 310.0 | 809. |
| 2 | -117.81 | 33.78 | 27.0 | 3589.0 | 507.0 | 1484. |
| 3 | -118.36 | 33.82 | 28.0 | 67.0 | 15.0 | 49. |
| 4 | -119.67 | 36.33 | 19.0 | 1241.0 | 244.0 | 850. |

# 데이터프레임에서 앞/뒤에 있는 5개 행만 출력하기

- df.head()
- df.head(n)
- df.tail()
- df.tail(n)

```
df=pd.read_csv('gapminder.tsv',sep='Wt')
```

```
df.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 |

gapminder 데이터는 세계 각 국가의 여러 지표에 대한 시계열 데이터를 포함하는 데이터셋입니다. 이 데이터셋은 1952년부터 2007년까지의 기간 동안 다음과 같은 지표를 포함하고 있습니다.

국가 (Country)
대륙 (Continent)
연도 (Year)
기대 수명 (Life Expectancy)
인구 (Population)
인당 GDP (GDP per Capita)

# 데이터프레임의 정보 확인

- 전체적인 정보 : df.info()
- 행과 열의 개수 : df.shape
- 열 이름 확인 : df.columns
- 각 열의 자료형 : df.dtypes
- 기초 통계 정보 : df.describe()

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1704 entries, 0 to 1703
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype  
---  --------     -----------------  -------
 0   country    1704 non-null   object 
 1   continent  1704 non-null   object 
 2   year       1704 non-null   int64  
 3   lifeExp    1704 non-null   float64
 4   pop        1704 non-null   int64  
 5   gdpPercap  1704  non-null  float64
dtypes: float64(2), int64(2), object(2)
memory usage: 80.0+ KB
```

df.shape

```
(1704, 6)
```

df.columns

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

df.dtypes

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

```
df.describe()
```

| | year | lifeExp | pop | gdpPercap |
|---|---|---|---|---|
| count | 1704.00000 | 1704.000000 | 1.704000e+03 | 1704.000000 |
| mean | 1979.50000 | 59.474439 | 2.960121e+07 | 7215.327081 |
| std | 17.26533 | 12.917107 | 1.061579e+08 | 9857.454543 |
| min | 1952.00000 | 23.599000 | 6.001100e+04 | 241.165876 |
| 25% | 1965.75000 | 48.198000 | 2.793664e+06 | 1202.060309 |
| 50% | 1979.50000 | 60.712500 | 7.023596e+06 | 3531.846988 |
| 75% | 1993.25000 | 70.845500 | 1.958522e+07 | 9325.462346 |
| max | 2007.00000 | 82.603000 | 1.318683e+09 | 113523.132900 |

# 데이터 선택하기(인덱싱과 슬라이싱)

- 열만 택하기
  - df.열이름
  - df['열이름']
  - df[['열이름',...]] <---열이 2개 이상이면 데이터프레임으로 처리해야 함. 행만 선택하기
  - df[ ] : 반드시 연속된 행만 선택 가능하다. 불연속적인 경우 loc나 iloc사용할 것.

```
df.country
```

```
0       Afghanistan
1       Afghanistan
2       Afghanistan
3       Afghanistan
4       Afghanistan
           ...
Name: country, Length: 1704, dtype: object
```

```
df['pop']
```

```
0        8425333
1        9240934
2       10267083
3       11537966
4       13079460
          ...
Name: pop, Length: 1704, dtype: int64
```

```
df[['country','continent','pop']]
```

| | country | continent | pop |
|---|---|---|---|
| 0 | Afghanistan | Asia | 8425333 |
| 1 | Afghanistan | Asia | 9240934 |
| 2 | Afghanistan | Asia | 10267083 |
| 3 | Afghanistan | Asia | 11537966 |
| 4 | Afghanistan | Asia | 13079460 |
| ... | ... | ... | ... |

1704 rows × 3 columns

```
df[10:13]
```

| | country | continent | year | lifeExp | pop | gdpPercap |
|---|---|---|---|---|---|---|
| 10 | Afghanistan | Asia | 2002 | 42.129 | 25268405 | 726.734055 |
| 11 | Afghanistan | Asia | 2007 | 43.828 | 31889923 | 974.580338 |
| 12 | Albania | Europe | 1952 | 55.230 | 1282697 | 1601.056136 |

# 데이터 선택하기(인덱싱과 슬라이싱)

- loc : 레이블 기반으로 선택. 행 번호는 생략 불가. 열은 반드시 열 이름을 입력해야 한다.
  - df.loc[행 번호, 열 이름]
  - df.loc[행 인덱스, 열 이름]

```
df.loc[0]
```

```
country      Afghanistan
continent           Asia
year                1952
lifeExp           28.801
pop              8425333
gdpPercap     779.445314
Name: 0, dtype: object
```

```
df.loc[1:5,['country','year','pop']]
```

| | country | year | pop |
|---|---|---|---|
| 1 | Afghanistan | 1957 | 9240934 |
| 2 | Afghanistan | 1962 | 10267083 |
| 3 | Afghanistan | 1967 | 11537966 |
| 4 | Afghanistan | 1972 | 13079460 |
| 5 | Afghanistan | 1977 | 14880372 |

# 데이터 선택하기(인덱싱과 슬라이싱)

- iloc : 행과 열의 번호를 지정해서 선택. 열 번호는 생략 가능.
  - df.iloc[행 번호, 열 번호]

```
df.iloc[1,0]
```

```
'Afghanistan'
```

```
df.iloc[:,[0,2,5]]
```

| | country | year | gdpPercap |
|---|---|---|---|
| 0 | Afghanistan | 1952 | 779.445314 |
| 1 | Afghanistan | 1957 | 820.853030 |
| 2 | Afghanistan | 1962 | 853.100710 |
| 3 | Afghanistan | 1967 | 836.197138 |
| 4 | Afghanistan | 1972 | 739.981106 |
| ... | ... | ... | ... |
| 1699 | Zimbabwe | 1987 | 706.157306 |
| 1700 | Zimbabwe | 1992 | 693.420786 |
| 1701 | Zimbabwe | 1997 | 792.449960 |
| 1702 | Zimbabwe | 2002 | 672.038623 |
| 1703 | Zimbabwe | 2007 | 469.709298 |

1704 rows × 3 columns

```
df.iloc[[0,3,7]]
```

| | country | continent | year | lifeExp | pop | gdpPercap |
|---|---|---|---|---|---|---|
| 0 | Afghanistan | Asia | 1952 | 28.801 | 8425333 | 779.445314 |
| 3 | Afghanistan | Asia | 1967 | 34.020 | 11537966 | 836.197138 |
| 7 | Afghanistan | Asia | 1987 | 40.822 | 13867957 | 852.395945 |

```
df.iloc[::-2].head()
```

| | country | continent | year | lifeExp | pop | gdpPercap |
|---|---|---|---|---|---|---|
| 1703 | Zimbabwe | Africa | 2007 | 43.487 | 12311143 | 469.709298 |
| 1701 | Zimbabwe | Africa | 1997 | 46.809 | 11404948 | 792.449960 |
| 1699 | Zimbabwe | Africa | 1987 | 62.351 | 9216418 | 706.157306 |