

```
In [ ]: import pandas as pd
import warnings
warnings.filterwarnings('ignore')
```

```
In [ ]: # renombrar fuente de datos
gapminder = pd.read_csv(r'./gapminder_full.csv')

# mostrar encabezado de los datos
gapminder.head()
```

```
Out[ ]:
```

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

Parte 1. Explorando los datos

```
In [ ]: # mirar la forma de los datos
gapminder.shape
```

```
Out[ ]: (1704, 6)
```

```
In [ ]: # mirar el tipo de dato por variable
gapminder.dtypes
```

```
Out[ ]: country      object
year          int64
population     int64
continent      object
life_exp      float64
gdp_cap       float64
dtype: object
```

Preguntas sobre country

¿Cuántos países hay en 2007?

```
In [ ]: gapminder_2007 = gapminder.query('year == 2007')
gapminder_2007.country.nunique()
```

```
Out[ ]: 142
```

¿Cuántos países por continente?

```
In [ ]: gapminder_2007.groupby('continent').count()['country']
```

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

Preguntas sobre year

¿De qué año a qué año van los datos?

```
In [ ]: print(f'Año inicial: {gapminder.year.min()}')
        print(f'Año final: {gapminder.year.max()}')
```

Año inicial: 1952

Año final: 2007

¿Cuántos datos hay por año?

```
In [ ]: gapminder.groupby('year').count()['country']
```

```
Out[ ]: year
1952    142
1957    142
1962    142
1967    142
1972    142
1977    142
1982    142
1987    142
1992    142
1997    142
2002    142
2007    142
Name: country, dtype: int64
```

Preguntas sobre población

¿Cuál es la población mínima y máxima para el año 2007?

```
In [ ]: gapminder_2007.population.agg(['min', 'max']).apply('{:,.0f}'.format)
```

```
Out[ ]: min      199,579
        max    1,318,683,096
        Name: population, dtype: object
```

¿Cuál es la población mundial por año?

```
In [ ]: gapminder.groupby('year').sum()['population'].apply('{:,.0f}'.format)
```

```
Out[ ]: year
1952    2,406,957,150
1957    2,664,404,580
1962    2,899,782,974
1967    3,217,478,384
1972    3,576,977,158
1977    3,930,045,807
1982    4,289,436,840
1987    4,691,477,418
1992    5,110,710,260
1997    5,515,204,472
2002    5,886,977,579
2007    6,251,013,179
Name: population, dtype: object
```

¿Cuál es el país con mayor y menor población para 2007?

```
In [ ]: gapminder_2007[['country', 'population']].agg(['min', 'max'])
```

```
Out[ ]:   country  population
min  Afghanistan    199579
max   Zimbabwe 1318683096
```

¿Cuál es la población por continente para 2007?

```
In [ ]: gapminder_2007.groupby('continent').sum()['population'].apply('{:,.0f}'.format)
```

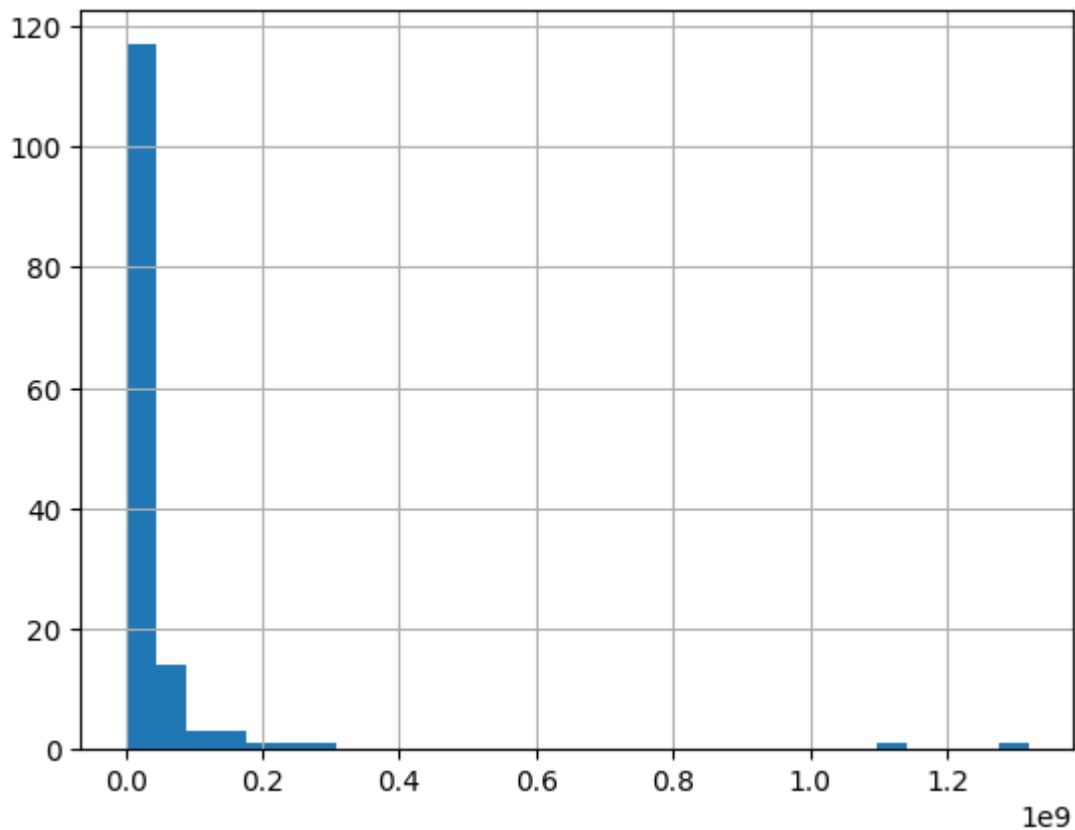
```
Out[ ]: continent
Africa          929,539,692
Americas        898,871,184
Asia            3,811,953,827
Europe          586,098,529
Oceania         24,549,947
Name: population, dtype: object
```

¿Cómo se distribuye (histograma) la población por año?

```
In [ ]: import matplotlib.pyplot as plt

gapminder_2007['population'].hist(bins=30)
```

```
Out[ ]: <Axes: >
```



Preguntas sobre continente

¿Cuáles continentes hay?

```
In [ ]: list(gapminder.continent.unique())
```

```
Out[ ]: ['Asia', 'Europe', 'Africa', 'Americas', 'Oceania']
```

¿Cuántos países hay en cada continente para 2007?

```
In [ ]: gapminder_2007.groupby('continent').count()['country']
```

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

Preguntas sobre experiencia de vida (life_exp)

¿Cuáles son los mínimos y los máximos de life_exp para el último año?

```
In [ ]: gapminder_2007['life_exp'].agg(['min', 'max'])
```

```
Out[ ]: min      39.613
max      82.603
Name: life_exp, dtype: float64
```

¿Cómo se distribuye el promedio de life_exp por continente para el último año?

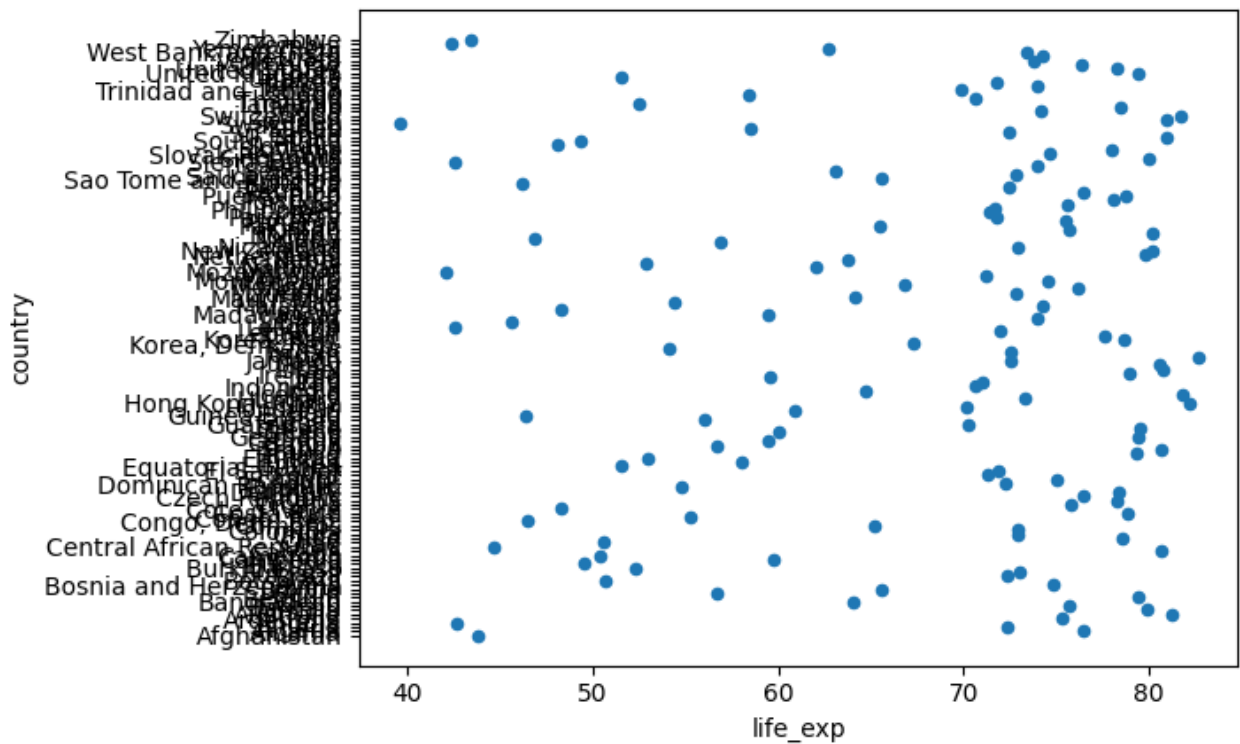
```
In [ ]: gapminder_2007.groupby('continent').mean()['life_exp']
```

```
Out[ ]: continent
Africa      54.806038
Americas    73.608120
Asia        70.728485
Europe      77.648600
Oceania     80.719500
Name: life_exp, dtype: float64
```

¿Cuál es el life_exp por país para 2007?

```
In [ ]: gapminder_2007[['country', 'life_exp']].plot.scatter(x='life_exp', y='country')
```

```
Out[ ]: <Axes: xlabel='life_exp', ylabel='country'>
```

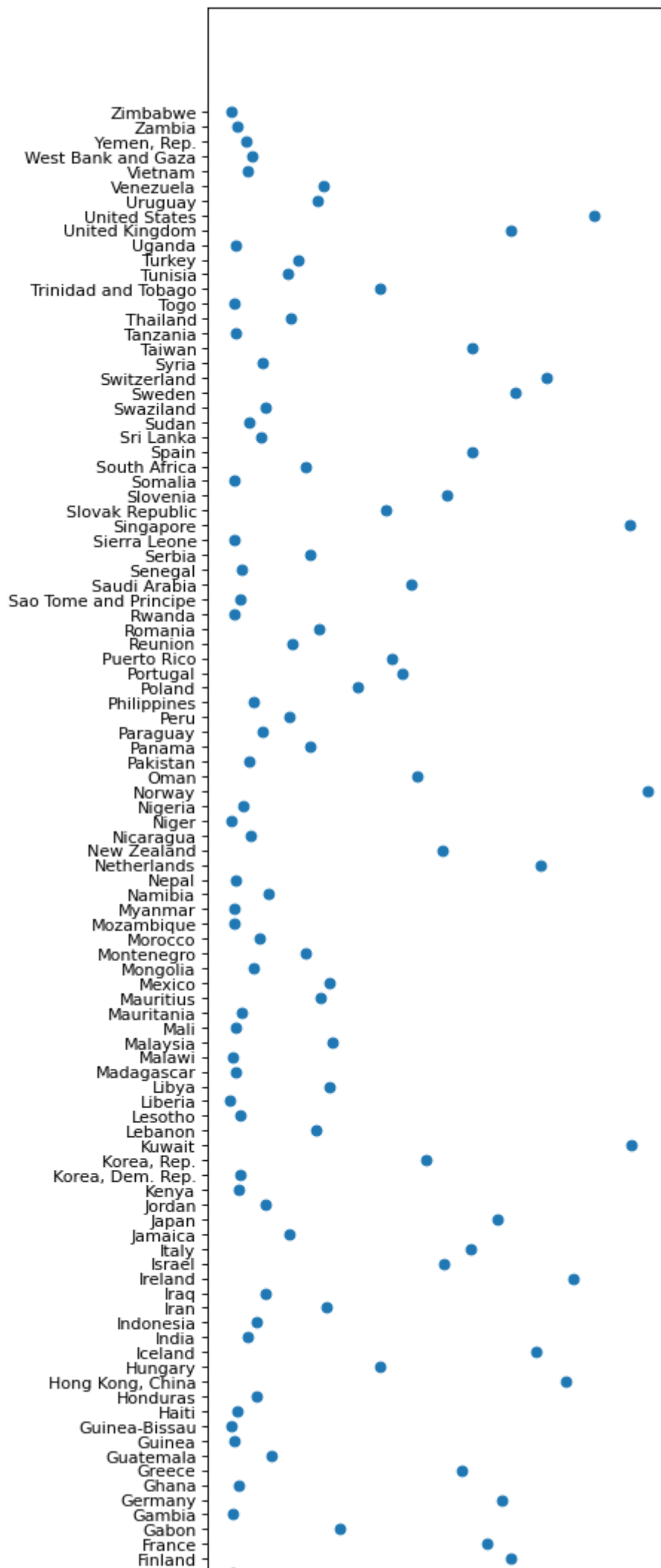


Preguntas sobre el límite del producto interno bruto (gdp_cap)

¿Cuál es el gdp_cap por país?

```
In [ ]: plt.figure(figsize=(5,25), dpi=80)
plt.scatter(x=gapminder_2007['gdp_cap'], y=gapminder_2007['country'])
```

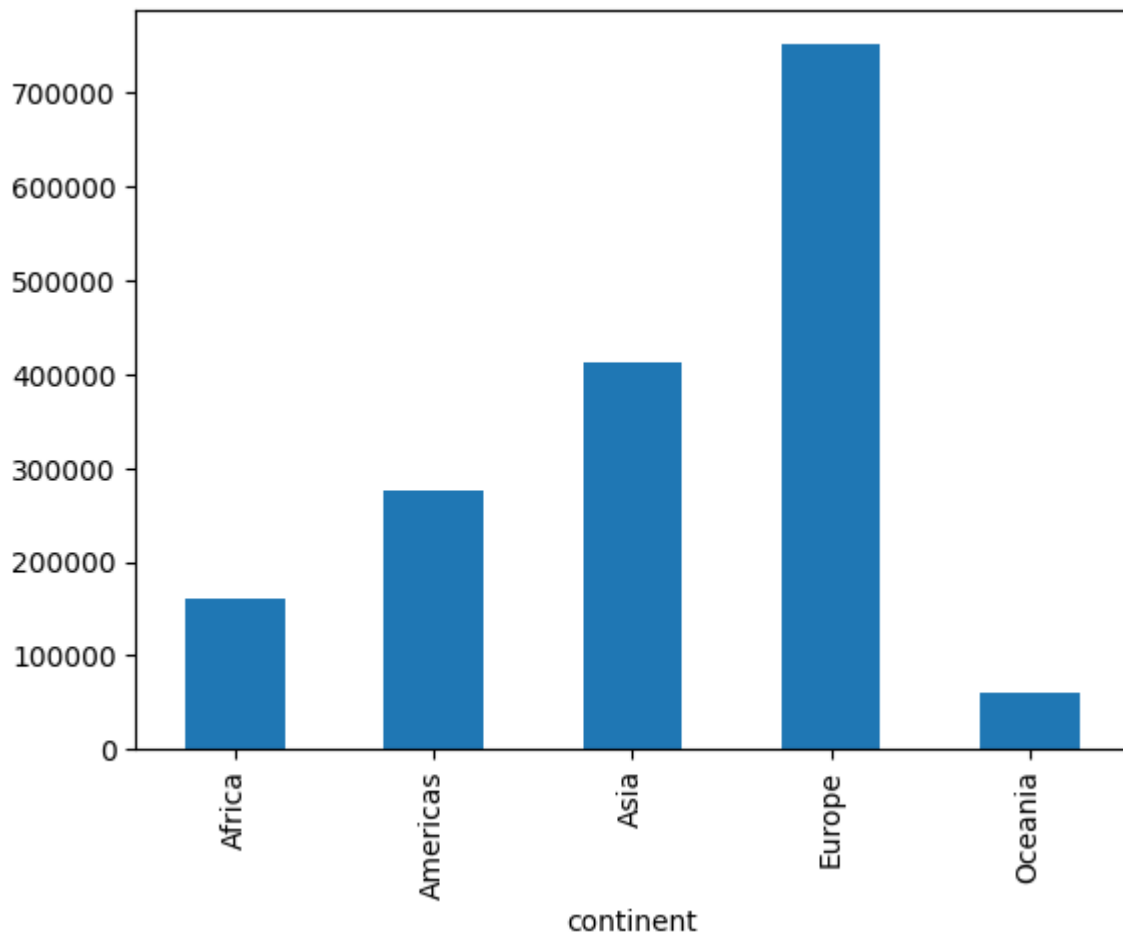
```
Out[ ]: <matplotlib.collections.PathCollection at 0x1590752cf90>
```



¿Cuál es el gdp_cap por continente?

```
In [ ]: gapminder_2007.groupby('continent').sum()['gdp_cap'].plot.bar()
```

```
Out[ ]: <Axes: xlabel='continent'>
```



¿Cómo se distribuye gdp per cap para el año 2007 (cuantiles)?

```
In [ ]: gapminder_2007['gdp_cap'].quantile([0.25,0.5,0.75]).apply('{:,.2f}'.format)
```

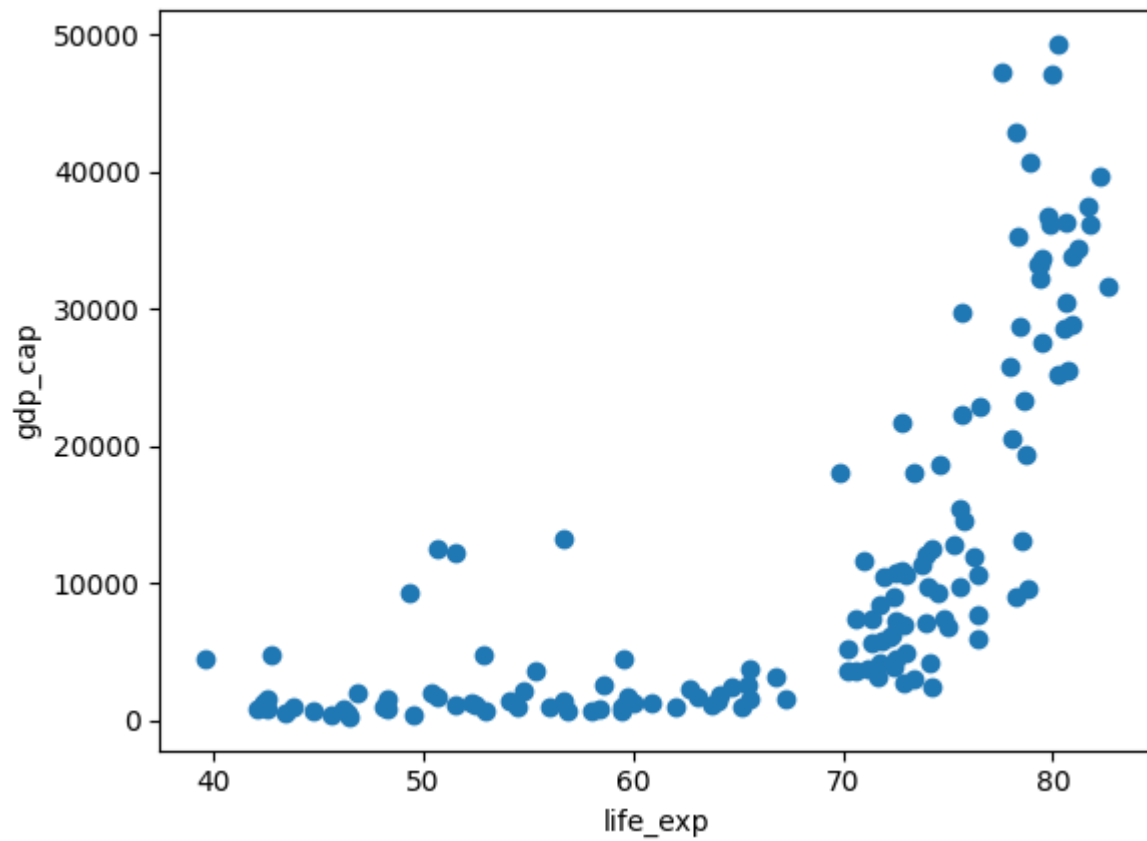
```
Out[ ]: 0.25    1,624.84
        0.50    6,124.37
        0.75   18,008.84
        Name: gdp_cap, dtype: object
```

Buscando relaciones entre las variables

¿A mayor life_exp, mayor gdp_cap?

```
In [ ]: plt.scatter(x=gapminder_2007['life_exp'], y=gapminder_2007['gdp_cap'])
        plt.xlabel('life_exp')
        plt.ylabel('gdp_cap')
```

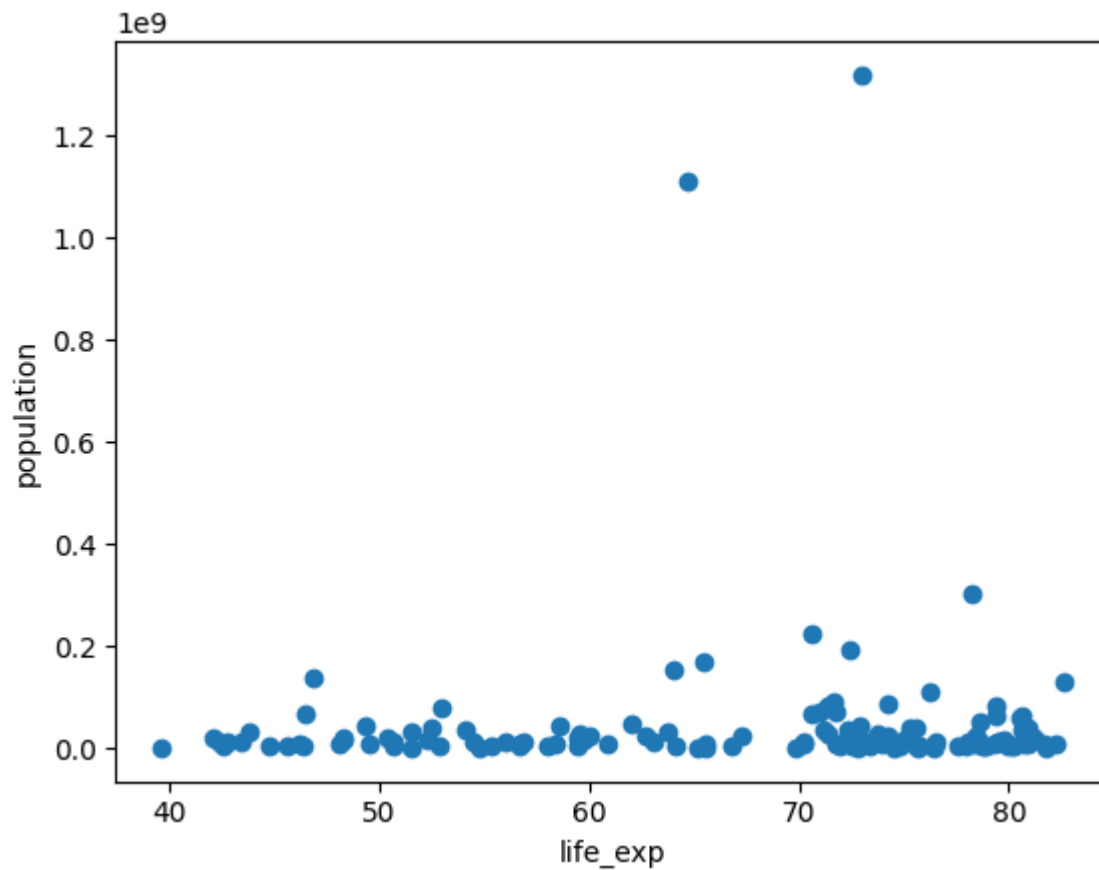
```
Out[ ]: Text(0, 0.5, 'gdp_cap')
```

¿A mayor población, mayor life_exp?

```
In [ ]: plt.scatter(x=gapminder_2007['life_exp'], y=gapminder_2007['population'])
plt.xlabel('life_exp')
plt.ylabel('population')
```

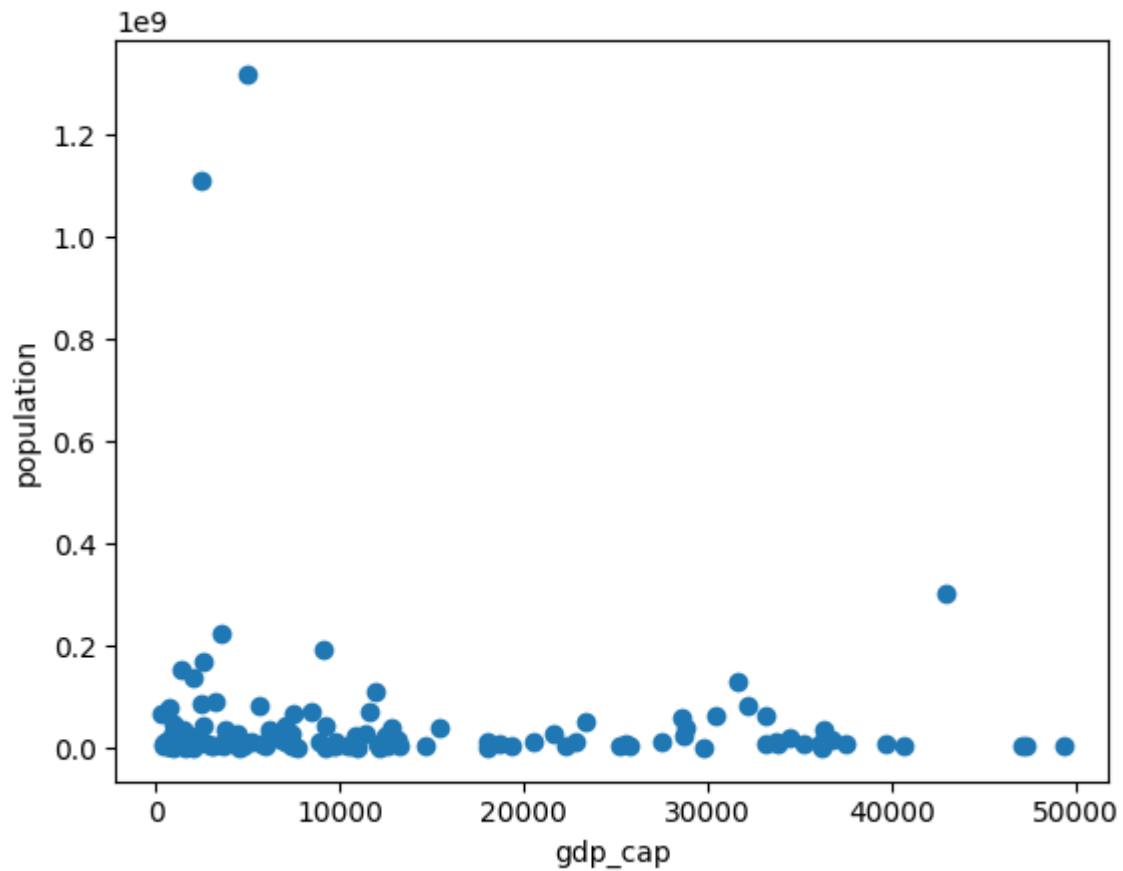
```
Out[ ]: Text(0, 0.5, 'population')
```



¿A mayor población, mayor gdp_cap?

```
In [ ]: plt.scatter(x=gapminder_2007['gdp_cap'], y=gapminder_2007['population'])  
plt.xlabel('gdp_cap')  
plt.ylabel('population')
```

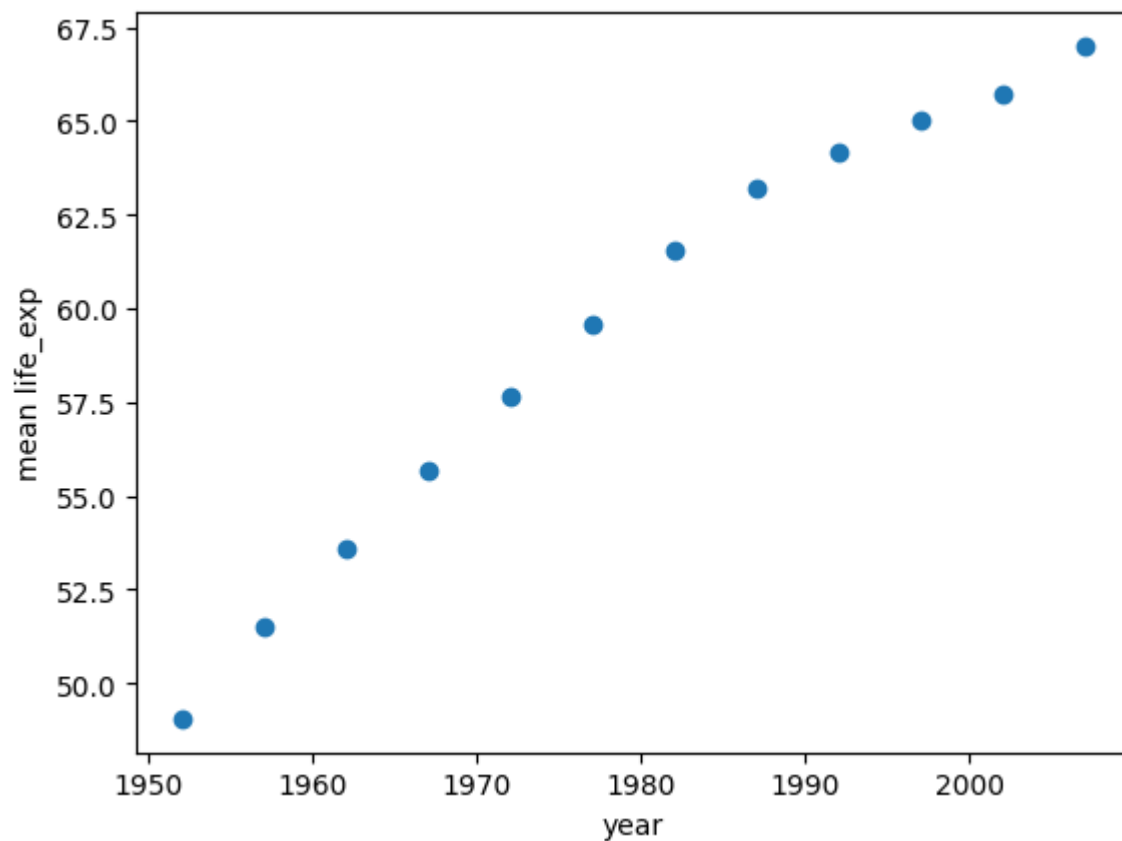
Out[]: Text(0, 0.5, 'population')



¿A mayor año, mayor life_exp?

```
In [ ]: lifeexp_year = pd.DataFrame(gapminder.groupby('year').mean()['life_exp'])  
plt.scatter(x=lifeexp_year.index, y=lifeexp_year.values)  
plt.xlabel('year')  
plt.ylabel('mean life_exp')
```

Out[]: Text(0, 0.5, 'mean life_exp')



¿A mayor año, mayor gdp_cap?

```
In [ ]: lifeexp_year = pd.DataFrame(gapminder.groupby('year').mean()['gdp_cap'])
plt.scatter(x=lifeexp_year.index, y=lifeexp_year.values)
plt.xlabel('year')
plt.ylabel('mean gdp_cap')
```

```
Out[ ]: Text(0, 0.5, 'mean gdp_cap')
```

