# SETUP

```
# !pip install hvplot
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import hvplot.pandas

from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.linear_model import LinearRegression

%matplotlib inline
```

```
led = pd.read_csv('/content/Life Expectancy Data.csv')
led
```

| | Country | Year | Status | Life expectancy | Adult Mortality | infant deaths | Alcohol | percentage expenditure | Hepatitis B | Measles | BMI | under-five deaths | Polio | Total expenditure | Dip |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Afghanistan | 2015 | Developing | 65.0 | 263.0 | 62 | 0.01 | 71.279624 | 65.0 | 1154 | 19.1 | 83 | 6.0 | 8.16 | |
| **1** | Afghanistan | 2014 | Developing | 59.9 | 271.0 | 64 | 0.01 | 73.523582 | 62.0 | 492 | 18.6 | 86 | 58.0 | 8.18 | |
| **2** | Afghanistan | 2013 | Developing | 59.9 | 268.0 | 66 | 0.01 | 73.219243 | 64.0 | 430 | 18.1 | 89 | 62.0 | 8.13 | |
| **3** | Afghanistan | 2012 | Developing | 59.5 | 272.0 | 69 | 0.01 | 78.184215 | 67.0 | 2787 | 17.6 | 93 | 67.0 | 8.52 | |
| **4** | Afghanistan | 2011 | Developing | 59.2 | 275.0 | 71 | 0.01 | 7.097109 | 68.0 | 3013 | 17.2 | 97 | 68.0 | 7.87 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **2933** | Zimbabwe | 2004 | Developing | 44.3 | 723.0 | 27 | 4.36 | 0.000000 | 68.0 | 31 | 27.1 | 42 | 67.0 | 7.13 | |
| **2934** | Zimbabwe | 2003 | Developing | 44.5 | 715.0 | 26 | 4.06 | 0.000000 | 7.0 | 998 | 26.7 | 41 | 7.0 | 6.52 | |
| **2935** | Zimbabwe | 2002 | Developing | 44.8 | 73.0 | 25 | 4.43 | 0.000000 | 73.0 | 304 | 26.3 | 40 | 73.0 | 6.53 | |
| **2936** | Zimbabwe | 2001 | Developing | 45.3 | 686.0 | 25 | 1.72 | 0.000000 | 76.0 | 529 | 25.9 | 39 | 76.0 | 6.16 | |
| **2937** | Zimbabwe | 2000 | Developing | 46.0 | 665.0 | 24 | 1.68 | 0.000000 | 79.0 | 1483 | 25.5 | 39 | 78.0 | 7.10 | |

2938 rows × 22 columns

```
led.shape
```

```
(2938, 22)
```

```
led.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2938 entries, 0 to 2937
Data columns (total 22 columns):
 #   Column                           Non-Null Count  Dtype  
---  ------                           --------------  -----  
 0   Country                          2938 non-null   object 
 1   Year                             2938 non-null   int64  
 2   Status                           2938 non-null   object 
 3   Life expectancy                  2928 non-null   float64
 4   Adult Mortality                  2928 non-null   float64
 5   infant deaths                    2938 non-null   int64  
 6   Alcohol                          2744 non-null   float64
 7   percentage expenditure           2938 non-null   float64
 8   Hepatitis B                      2385 non-null   float64
 9   Measles                          2938 non-null   int64  
 10   BMI                             2904 non-null   float64
 11  under-five deaths                2938 non-null   int64  
 12  Polio                            2919 non-null   float64
 13  Total expenditure                2712 non-null   float64
 14  Diphtheria                       2919 non-null   float64
 15   HIV/AIDS                        2938 non-null   float64
 16  GDP                              2490 non-null   float64
 17  Population                       2286 non-null   float64
 18   thinness  1-19 years            2904 non-null   float64
 19   thinness 5-9 years             2904 non-null   float64
 20  Income composition of resources  2771 non-null   float64
 21  Schooling                        2775 non-null   float64
dtypes: float64(16), int64(4), object(2)
memory usage: 505.1+ KB
```

```
led.drop('Status', axis=1, inplace=True)
```

```
led
```

| | Country | Year | Life expectancy | Adult Mortality | infant deaths | Alcohol | percentage expenditure | Hepatitis B | Measles | BMI | under-five deaths | Polio | Total expenditure | Diphtheria | HI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Afghanistan | 2015 | 65.0 | 263.0 | 62 | 0.01 | 71.279624 | 65.0 | 1154 | 19.1 | 83 | 6.0 | 8.16 | 65.0 | |
| **1** | Afghanistan | 2014 | 59.9 | 271.0 | 64 | 0.01 | 73.523582 | 62.0 | 492 | 18.6 | 86 | 58.0 | 8.18 | 62.0 | |
| **2** | Afghanistan | 2013 | 59.9 | 268.0 | 66 | 0.01 | 73.219243 | 64.0 | 430 | 18.1 | 89 | 62.0 | 8.13 | 64.0 | |
| **3** | Afghanistan | 2012 | 59.5 | 272.0 | 69 | 0.01 | 78.184215 | 67.0 | 2787 | 17.6 | 93 | 67.0 | 8.52 | 67.0 | |
| **4** | Afghanistan | 2011 | 59.2 | 275.0 | 71 | 0.01 | 7.097109 | 68.0 | 3013 | 17.2 | 97 | 68.0 | 7.87 | 68.0 | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **2933** | Zimbabwe | 2004 | 44.3 | 723.0 | 27 | 4.36 | 0.000000 | 68.0 | 31 | 27.1 | 42 | 67.0 | 7.13 | 65.0 | |
| **2934** | Zimbabwe | 2003 | 44.5 | 715.0 | 26 | 4.06 | 0.000000 | 7.0 | 998 | 26.7 | 41 | 7.0 | 6.52 | 68.0 | |
| **2935** | Zimbabwe | 2002 | 44.8 | 73.0 | 25 | 4.43 | 0.000000 | 73.0 | 304 | 26.3 | 40 | 73.0 | 6.53 | 71.0 | |
| **2936** | Zimbabwe | 2001 | 45.3 | 686.0 | 25 | 1.72 | 0.000000 | 76.0 | 529 | 25.9 | 39 | 76.0 | 6.16 | 75.0 | |
| **2937** | Zimbabwe | 2000 | 46.0 | 665.0 | 24 | 1.68 | 0.000000 | 79.0 | 1483 | 25.5 | 39 | 78.0 | 7.10 | 78.0 | |

2938 rows × 21 columns

```
1 bcled = led.groupby('Country').mean()
```

```
1 bcled
```

| Country | Year | Life expectancy | Adult Mortality | infant deaths | Alcohol | percentage expenditure | Hepatitis B | Measles | BMI | under-five deaths | Polio | Total expenditure | Diphther |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Afghanistan** | 2007.5 | 58.19375 | 269.0625 | 78.2500 | 0.014375 | 34.960110 | 64.562500 | 2362.2500 | 15.51875 | 107.5625 | 48.3750 | 8.252500 | 52.3 |
| **Albania** | 2007.5 | 75.15625 | 45.0625 | 0.6875 | 4.848750 | 193.259091 | 98.000000 | 53.3750 | 49.06875 | 0.9375 | 98.1250 | 5.945625 | 98.0 |
| **Algeria** | 2007.5 | 73.61875 | 108.1875 | 20.3125 | 0.406667 | 236.185241 | 78.000000 | 1943.8750 | 48.74375 | 23.5000 | 91.7500 | 4.604000 | 91.8 |
| **Angola** | 2007.5 | 49.01875 | 328.5625 | 83.7500 | 5.740667 | 102.100268 | 70.222222 | 3561.3125 | 18.01875 | 132.6250 | 46.1250 | 3.919333 | 47.6 |
| **Antigua and Barbuda** | 2007.5 | 75.05625 | 127.5000 | 0.0000 | 7.949333 | 1001.585226 | 98.266667 | 0.0000 | 38.42500 | 0.0000 | 96.9375 | 4.791333 | 98.3 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **Venezuela (Bolivarian Republic of)** | 2007.5 | 73.38750 | 163.0000 | 9.3750 | 7.420000 | 0.000000 | 66.250000 | 165.0000 | 54.48750 | 10.7500 | 74.6875 | 4.998667 | 68.5 |
| **Viet Nam** | 2007.5 | 74.77500 | 126.5625 | 29.1875 | 3.087333 | 0.000000 | 87.538462 | 4232.9375 | 11.18750 | 36.5000 | 94.9375 | 5.977333 | 91.7 |
| **Yemen** | 2007.5 | 63.86250 | 211.8125 | 39.3750 | 0.047333 | 0.000000 | 55.687500 | 2761.1875 | 33.48750 | 51.6250 | 67.1250 | 5.005333 | 72.6 |
| **Zambia** | 2007.5 | 53.90625 | 354.3125 | 33.4375 | 2.239333 | 89.650407 | 69.818182 | 6563.8125 | 17.45000 | 52.3750 | 64.3750 | 5.824000 | 74.2 |
| **Zimbabwe** | 2007.5 | 50.48750 | 462.3750 | 26.5625 | 4.482000 | 20.364271 | 70.562500 | 923.0000 | 25.13750 | 40.8125 | 75.6250 | 6.158667 | 75.1 |

193 rows × 20 columns

Next steps: View recommended plots

```
1 bcled.fillna(0)
```

| Country | Year | Life expectancy | Adult Mortality | infant deaths | Alcohol | percentage expenditure | Hepatitis B | Measles | BMI | under-five deaths | Polio | Total expenditure | Diphther |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Afghanistan** | 2007.5 | 58.19375 | 269.0625 | 78.2500 | 0.014375 | 34.960110 | 64.562500 | 2362.2500 | 15.51875 | 107.5625 | 48.3750 | 8.252500 | 52.3 |
| **Albania** | 2007.5 | 75.15625 | 45.0625 | 0.6875 | 4.848750 | 193.259091 | 98.000000 | 53.3750 | 49.06875 | 0.9375 | 98.1250 | 5.945625 | 98.0 |
| **Algeria** | 2007.5 | 73.61875 | 108.1875 | 20.3125 | 0.406667 | 236.185241 | 78.000000 | 1943.8750 | 48.74375 | 23.5000 | 91.7500 | 4.604000 | 91.8 |
| **Angola** | 2007.5 | 49.01875 | 328.5625 | 83.7500 | 5.740667 | 102.100268 | 70.222222 | 3561.3125 | 18.01875 | 132.6250 | 46.1250 | 3.919333 | 47.6 |
| **Antigua and Barbuda** | 2007.5 | 75.05625 | 127.5000 | 0.0000 | 7.949333 | 1001.585226 | 98.266667 | 0.0000 | 38.42500 | 0.0000 | 96.9375 | 4.791333 | 98.3 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **Venezuela (Bolivarian Republic of)** | 2007.5 | 73.38750 | 163.0000 | 9.3750 | 7.420000 | 0.000000 | 66.250000 | 165.0000 | 54.48750 | 10.7500 | 74.6875 | 4.998667 | 68.5 |
| **Viet Nam** | 2007.5 | 74.77500 | 126.5625 | 29.1875 | 3.087333 | 0.000000 | 87.538462 | 4232.9375 | 11.18750 | 36.5000 | 94.9375 | 5.977333 | 91.7 |
| **Yemen** | 2007.5 | 63.86250 | 211.8125 | 39.3750 | 0.047333 | 0.000000 | 55.687500 | 2761.1875 | 33.48750 | 51.6250 | 67.1250 | 5.005333 | 72.6 |
| **Zambia** | 2007.5 | 53.90625 | 354.3125 | 33.4375 | 2.239333 | 89.650407 | 69.818182 | 6563.8125 | 17.45000 | 52.3750 | 64.3750 | 5.824000 | 74.2 |
| **Zimbabwe** | 2007.5 | 50.48750 | 462.3750 | 26.5625 | 4.482000 | 20.364271 | 70.562500 | 923.0000 | 25.13750 | 40.8125 | 75.6250 | 6.158667 | 75.1 |

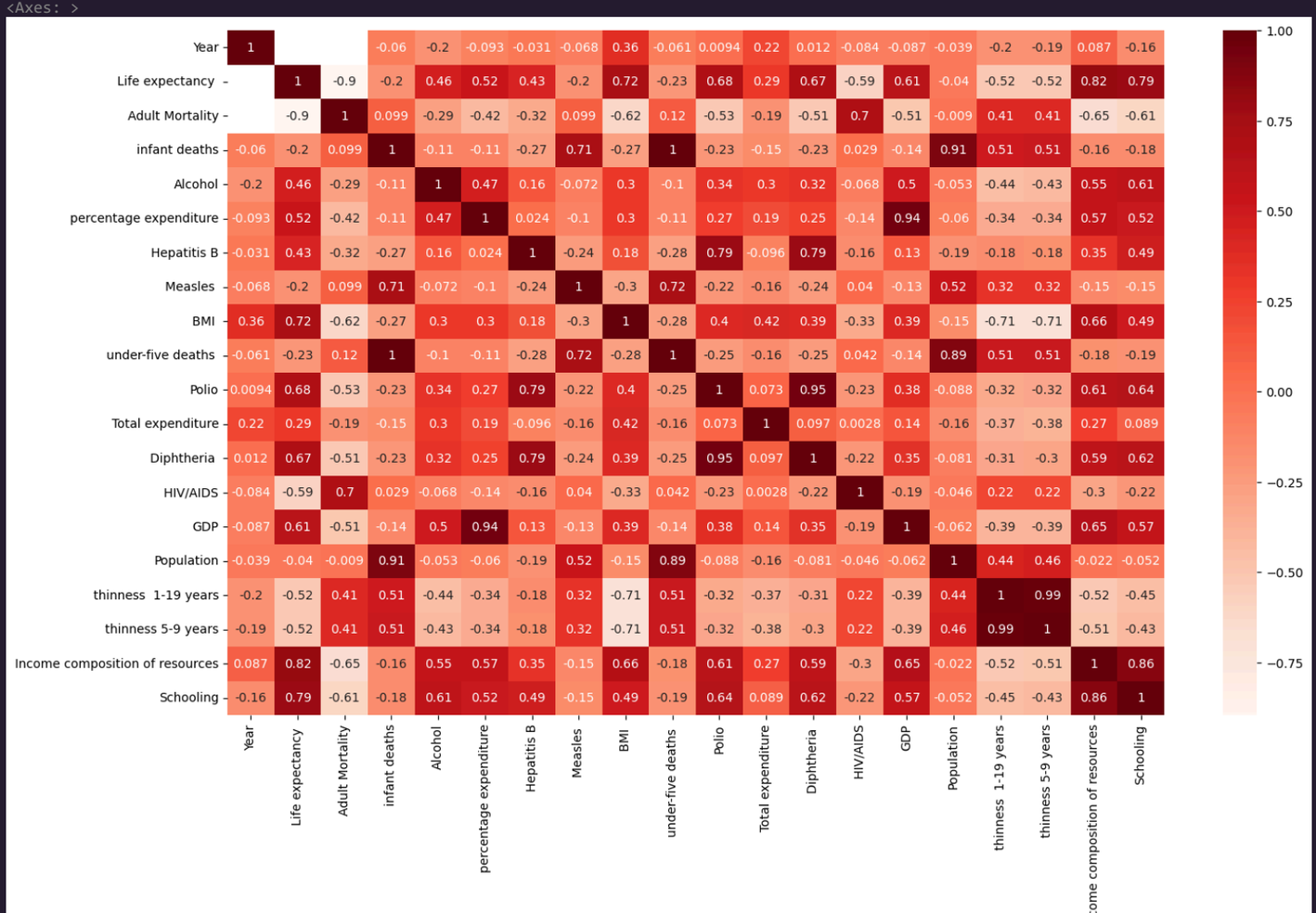
193 rows × 20 columns

```
1 bcled.corr().fillna(0)
```

| | Year | Life expectancy | Adult Mortality | infant deaths | Alcohol | percentage expenditure | Hepatitis B | Measles | BMI | under-five deaths | Polio | Total expenditure |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Year** | 1.000000 | 0.000000 | 0.000000 | -0.059727 | -0.200113 | -0.093238 | -0.030887 | -0.068415 | 0.357401 | -0.061041 | 0.009390 | 0.221568 |
| **Life expectancy** | 0.000000 | 1.000000 | -0.896441 | -0.199914 | 0.461720 | 0.524320 | 0.429536 | -0.201076 | 0.723824 | -0.225785 | 0.679231 | 0.290713 |
| **Adult Mortality** | 0.000000 | -0.896441 | 1.000000 | 0.099440 | -0.286503 | -0.424081 | -0.321612 | 0.098834 | -0.621810 | 0.119620 | -0.531039 | -0.191233 |
| **infant deaths** | -0.059727 | -0.199914 | 0.099440 | 1.000000 | -0.107259 | -0.111340 | -0.267790 | 0.712231 | -0.272189 | 0.996969 | -0.225164 | -0.154177 |
| **Alcohol** | -0.200113 | 0.461720 | -0.286503 | -0.107259 | 1.000000 | 0.465545 | 0.160581 | -0.072284 | 0.299362 | -0.104307 | 0.339619 | 0.301748 |
| **percentage expenditure** | -0.093238 | 0.524320 | -0.424081 | -0.111340 | 0.465545 | 1.000000 | 0.023746 | -0.103751 | 0.295305 | -0.114447 | 0.267732 | 0.193936 |
| **Hepatitis B** | -0.030887 | 0.429536 | -0.321612 | -0.267790 | 0.160581 | 0.023746 | 1.000000 | -0.239001 | 0.182466 | -0.280908 | 0.787961 | -0.096168 |
| **Measles** | -0.068415 | -0.201076 | 0.098834 | 0.712231 | -0.072284 | -0.103751 | -0.239001 | 1.000000 | -0.295474 | 0.720787 | -0.223534 | -0.158464 |
| **BMI** | 0.357401 | 0.723824 | -0.621810 | -0.272189 | 0.299362 | 0.295305 | 0.182466 | -0.295474 | 1.000000 | -0.284539 | 0.398468 | 0.418922 |
| **under-five deaths** | -0.061041 | -0.225785 | 0.119620 | 0.996969 | -0.104307 | -0.114447 | -0.280908 | 0.720787 | -0.284539 | 1.000000 | -0.249071 | -0.156143 |
| **Polio** | 0.009390 | 0.679231 | -0.531039 | -0.225164 | 0.339619 | 0.267732 | 0.787961 | -0.223534 | 0.398468 | -0.249071 | 1.000000 | 0.072834 |
| **Total expenditure** | 0.221568 | 0.290713 | -0.191233 | -0.154177 | 0.301748 | 0.193936 | -0.096168 | -0.158464 | 0.418922 | -0.156143 | 0.072834 | 1.000000 |
| **Diphtheria** | 0.011885 | 0.672322 | -0.511538 | -0.225175 | 0.324276 | 0.252658 | 0.788942 | -0.242165 | 0.390496 | -0.250265 | 0.952538 | 0.097214 |
| **HIV/AIDS** | -0.084259 | -0.587153 | 0.696999 | 0.029490 | -0.067714 | -0.138381 | -0.162508 | 0.040098 | -0.327423 | 0.042484 | -0.228528 | 0.002766 |
| **GDP** | -0.087401 | 0.611808 | -0.505445 | -0.138566 | 0.497086 | 0.942144 | 0.133620 | -0.133371 | 0.387616 | -0.143792 | 0.375788 | 0.137550 |
| **Population** | -0.039087 | -0.039915 | -0.009048 | 0.906180 | -0.052627 | -0.060158 | -0.193278 | 0.524840 | -0.154491 | 0.891157 | -0.087765 | -0.160807 |
| **thinness 1-19 years** | -0.195299 | -0.523989 | 0.412011 | 0.510734 | -0.444065 | -0.341189 | -0.176110 | 0.320336 | -0.707251 | 0.512663 | -0.324251 | -0.372911 |
| **thinness 5-9 years** | -0.194623 | -0.515970 | 0.414751 | 0.513618 | -0.430497 | -0.343198 | -0.183453 | 0.319666 | -0.708735 | 0.514090 | -0.316699 | -0.381227 |
| **Income composition of resources** | 0.086792 | 0.817545 | -0.651998 | -0.162852 | 0.551866 | 0.572658 | 0.346731 | -0.148016 | 0.661409 | -0.182633 | 0.611250 | 0.266547 |
| **Schooling** | -0.157704 | 0.794457 | -0.606786 | -0.176990 | 0.605842 | 0.515125 | 0.488121 | -0.148674 | 0.489900 | -0.191605 | 0.643011 | 0.088646 |

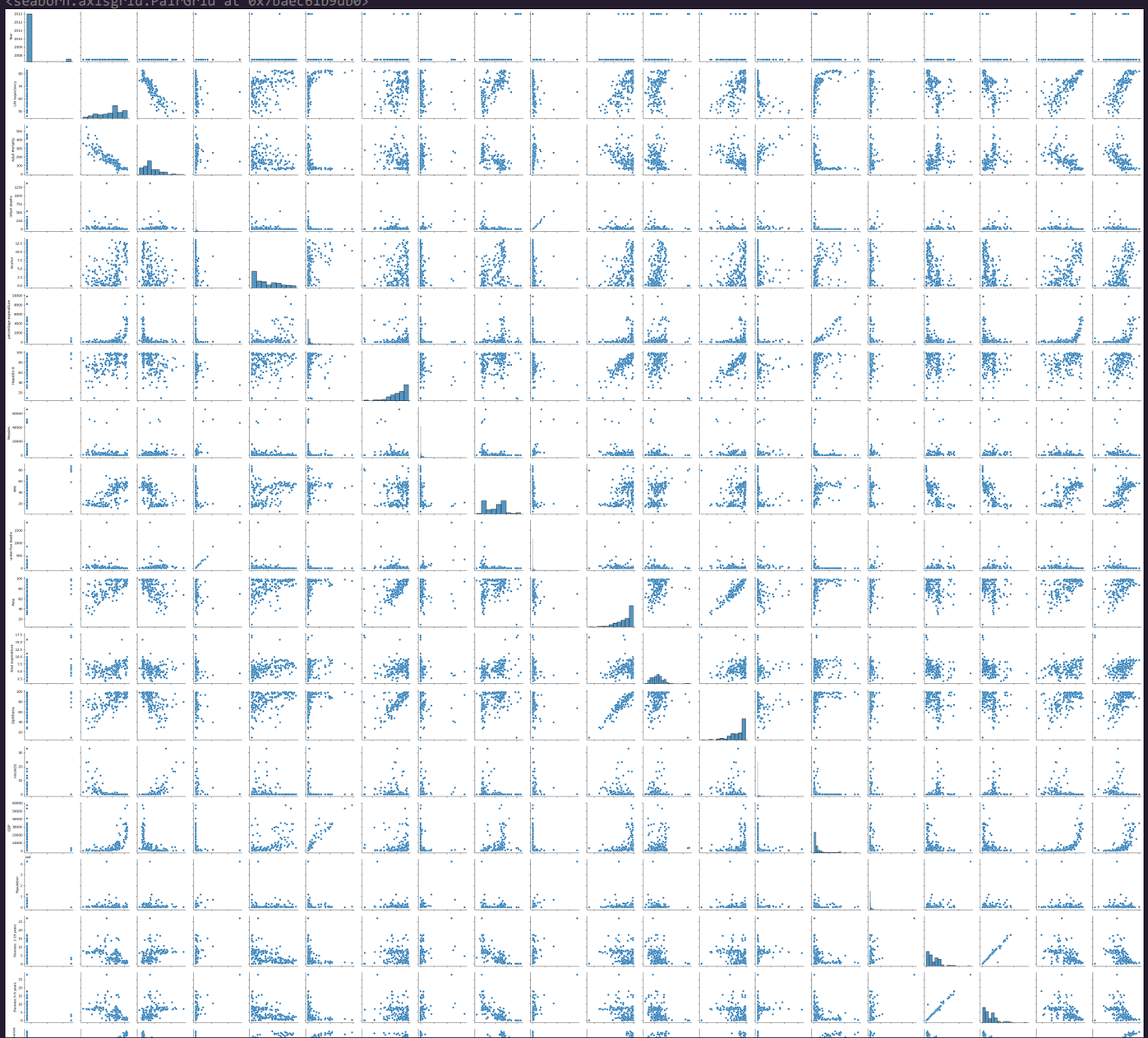
```
1 plt.figure(figsize=(17, 10))
2 heatmap = sns.heatmap(bcled.corr(), annot=True, cmap='Reds')
3 heatmap
```

```
<Axes: >
```




```
1 sns.pairplot(bcled)
```

```
<seaborn.axisgrid.PairGrid at 0x7baec61b9db0>
```

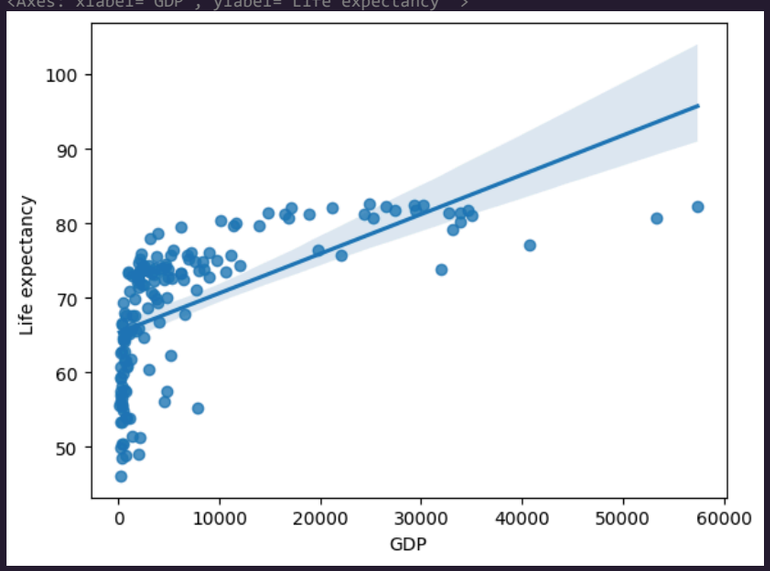


```
1 bcled.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 193 entries, Afghanistan to Zimbabwe
Data columns (total 20 columns):
 #   Column                           Non-Null Count  Dtype  
---  ------                           --------------  -----  
 0   Year                             193 non-null    float64
 1   Life expectancy                  183 non-null    float64
 2   Adult Mortality                  183 non-null    float64
 3   infant deaths                    193 non-null    float64
 4   Alcohol                          191 non-null    float64
 5   percentage expenditure           193 non-null    float64
 6   Hepatitis B                      184 non-null    float64
 7   Measles                          193 non-null    float64
 8    BMI                             189 non-null    float64
 9   under-five deaths                193 non-null    float64
 10  Polio                            193 non-null    float64
 11  Total expenditure                191 non-null    float64
 12  Diphtheria                       193 non-null    float64
 13   HIV/AIDS                        193 non-null    float64
 14  GDP                              163 non-null    float64
 15  Population                       145 non-null    float64
 16   thinness  1-19 years            189 non-null    float64
 17   thinness 5-9 years              189 non-null    float64
 18  Income composition of resources  176 non-null    float64
 19  Schooling                        180 non-null    float64
dtypes: float64(20)
memory usage: 35.7+ KB
```

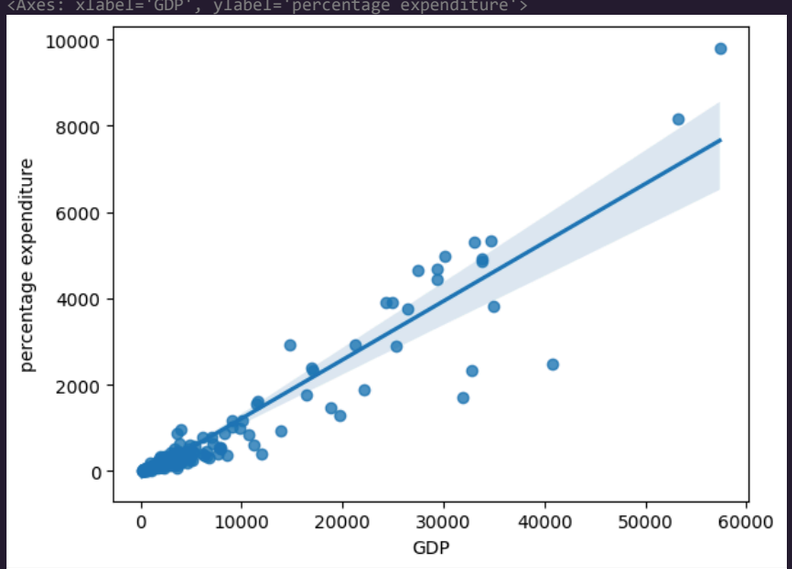
```
1 sns.regplot(x=bcled['GDP'], y=bcled['Life expectancy '])
```

```
<Axes: xlabel='GDP', ylabel='Life expectancy '>
```




```
1 sns.regplot(x=bcled['GDP'], y=bcled['percentage expenditure'])
```

```
<Axes: xlabel='GDP', ylabel='percentage expenditure'>
```




```
1 sns.regplot(x=bcled['Population'], y=bcled['under-five deaths '])
```

```
<Axes: xlabel='Population', ylabel='under-five deaths '>
```