In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import linear_model
```

In [3]:
```
df = pd.read_csv('C:/Users/nisho/Documents/SEM 5/ML and core applications/Life Expectancy Data.csv')
df.head()
```

Out[3]:

	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	...	Polio	Total expenditure	Diphtheria
0	Afghanistan	2015	Developing	65.0	263.0	62	0.01	71.279624	65.0	1154	...	6.0	8.16	65.0
1	Afghanistan	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0	492	...	58.0	8.18	62.0
2	Afghanistan	2013	Developing	59.9	268.0	66	0.01	73.219243	64.0	430	...	62.0	8.13	64.0
3	Afghanistan	2012	Developing	59.5	272.0	69	0.01	78.184215	67.0	2787	...	67.0	8.52	67.0
4	Afghanistan	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0	3013	...	68.0	7.87	68.0

rows × 22 columns

In [4]:
```
df.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
```

In [5]:
```
df.drop(['Country', 'Status'], axis=1, inplace=True)
df.head()
```

Out[5]:

	Year	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	Polio	Total expenditure	Diphtheria	HIV/AIDS
0	2015	65.0	263.0	62	0.01	71.279624	65.0	1154	19.1	83	6.0	8.16	65.0	0.1
1	2014	59.9	271.0	64	0.01	73.523582	62.0	492	18.6	86	58.0	8.18	62.0	0.1
2	2013	59.9	268.0	66	0.01	73.219243	64.0	430	18.1	89	62.0	8.13	64.0	0.1
3	2012	59.5	272.0	69	0.01	78.184215	67.0	2787	17.6	93	67.0	8.52	67.0	0.1
4	2011	59.2	275.0	71	0.01	7.097109	68.0	3013	17.2	97	68.0	7.87	68.0	0.1

In [6]:
```
df.isnull().values.any()
```

Out[6]:
```
True
```

In [7]:
```
df.isna().any()
```

Out[7]:
```
Year                               False
Life expectancy                     True
Adult Mortality                     True
infant deaths                      False
Alcohol                             True
percentage expenditure             False
Hepatitis B                         True
Measles                            False
 BMI                                True
under-five deaths                  False
Polio                               True
Total expenditure                   True
Diphtheria                          True
 HIV/AIDS                          False
GDP                                 True
Population                          True
 thinness  1-19 years               True
 thinness 5-9 years                 True
Income composition of resources     True
Schooling                           True
dtype: bool
```

In [8]:
```
df['Adult Mortality'].fillna(df['Adult Mortality'].median(), inplace=True)
```

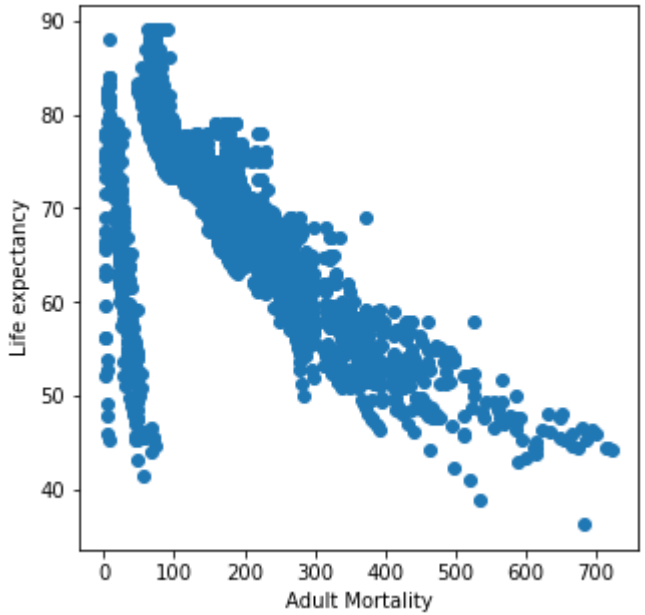
In [9]:
```
df['Life expectancy '].fillna(df['Life expectancy '].median(), inplace=True)
```

In [10]:
```
df['Hepatitis B'].fillna(df['Hepatitis B'].median(), inplace=True)
```

In [11]:
```
df['Polio'].fillna(df['Polio'].median(), inplace=True)
```

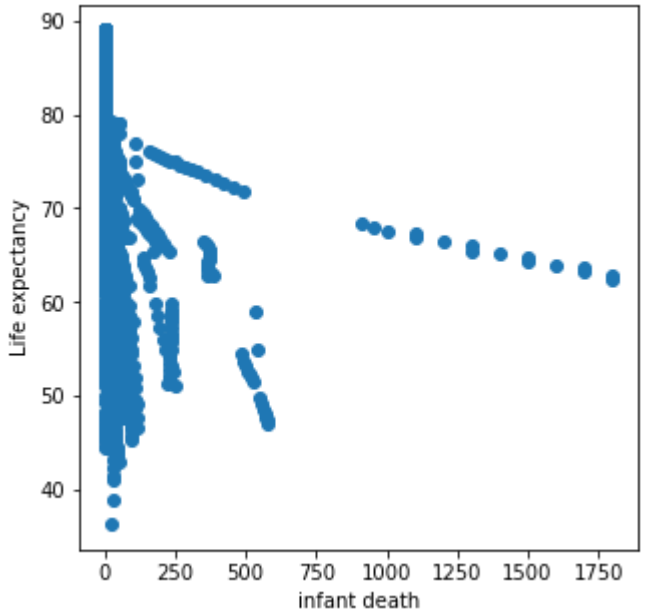
In [12]:
```
plt.figure(figsize=(5,5))
plt.xlabel('Adult Mortality')
plt.ylabel('Life expectancy')
plt.scatter(df['Adult Mortality'], df['Life expectancy '])
```

Out[12]:
```
<matplotlib.collections.PathCollection at 0x27fb747f310>
```



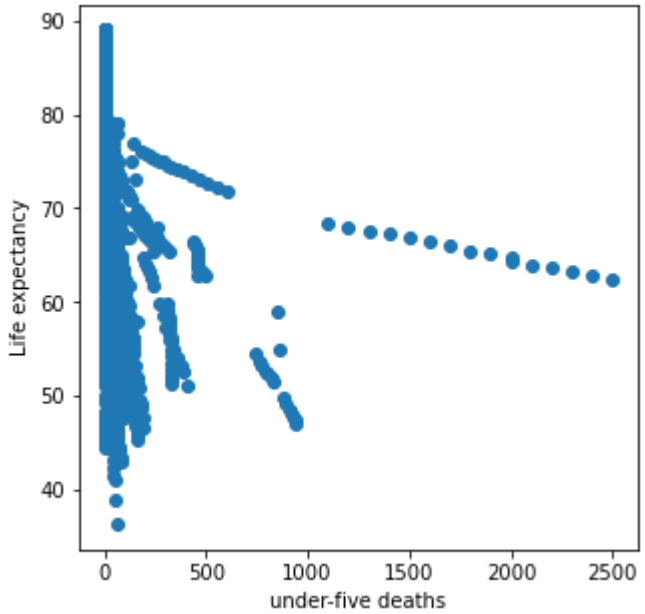
In [13]:
```
plt.figure(figsize=(5,5))
plt.xlabel('infant death')
plt.ylabel('Life expectancy')
plt.scatter(df['infant deaths'], df['Life expectancy '])
```

Out[13]:
```
<matplotlib.collections.PathCollection at 0x27fb75b92b0>
```



In [14]:
```
plt.figure(figsize=(5,5))
plt.xlabel('under-five deaths')
plt.ylabel('Life expectancy')
plt.scatter(df['under-five deaths '], df['Life expectancy '])
```

Out[14]:
```
<matplotlib.collections.PathCollection at 0x27fb761e8e0>
```



In [15]:
```
del = linear_model.LinearRegression().fit(df[['Adult Mortality', 'infant deaths', 'under-five deaths ']], df['Li
```

In [16]:
```
model.coef_
```

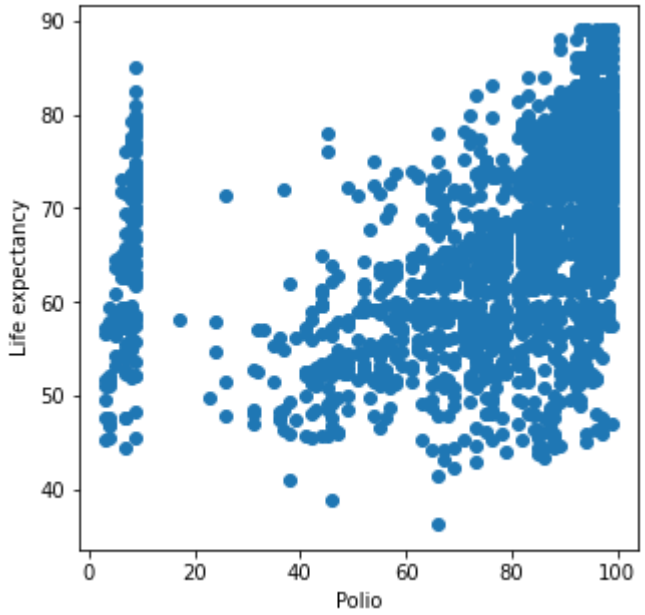
Out[16]:
```
array([-0.04954221,  0.18538215, -0.14538407])
```

In [18]:
```
model.intercept_
```

Out[18]:
```
77.88910698885105
```

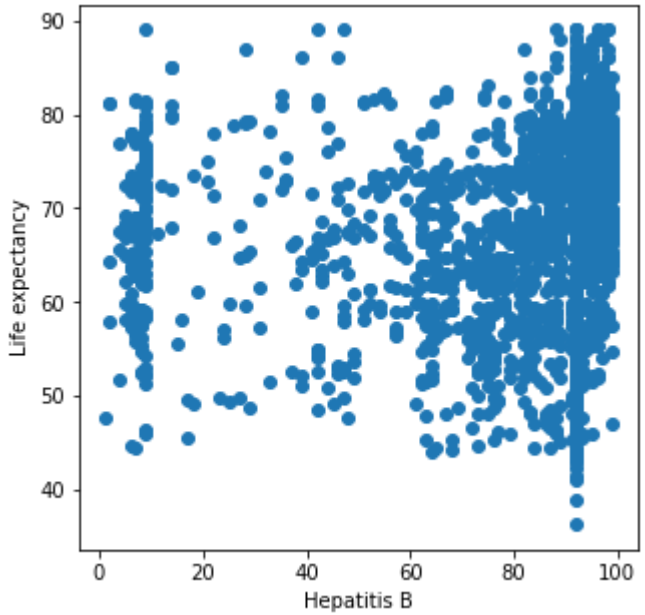
In [19]:
```
plt.figure(figsize=(5,5))
plt.xlabel('Polio')
plt.ylabel('Life expectancy')
plt.scatter(df['Polio'], df['Life expectancy '])
```

Out[19]:
```
<matplotlib.collections.PathCollection at 0x1d01077a4c0>
```



In [20]:
```
plt.figure(figsize=(5,5))
plt.xlabel('Hepatitis B')
plt.ylabel('Life expectancy')
plt.scatter(df['Hepatitis B'], df['Life expectancy '])
```

Out[20]: <matplotlib.collections.PathCollection at 0x1d010e4f580>



In [21]:
```
rRegression().fit(df[['Adult Mortality', 'infant deaths', 'under-five deaths ', 'Polio', 'Hepatitis B']], df['Li
```

In [22]:
```
model2.coef_
```

Out[22]:
```
array([-0.04540593,  0.14347081, -0.11237082,  0.10245096, -0.0083418 ])
```

In [23]:
```
model2.intercept_
```

Out[23]:
```
69.31838278076285
```

In [19]:
```
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import KFold


cv = float(format(cross_val_score(model,df[['Adult Mortality','infant deaths', 'under-five deaths ', ]],df['Life
```

In [20]:
```
print('Intercept: {}'.format(model.intercept_))
print('Coefficient: {}'.format(model.coef_))
```

```
Intercept: 77.88910698885105
Coefficient: [-0.04954221  0.18538215 -0.14538407]
```

In [23]:
```
print (cv)
```

```
0.468
```

In [ ]: