```
import numpy as np
import pandas as pd
from datascience import *
import matplotlib.pyplot as plt
import seaborn as sns
import random
import scipy.stats as stats
%matplotlib inline
plt.style.use('fivethirtyeight')
from google.colab import files
```

```
files.upload()
```

- **country_profile_variables.csv**(text/csv) - 9104 bytes, last modified: 10/26/2021 - 100% done

```
Saving country_profile_variables.csv to country_profile_variables.csv
{'country_profile_variables.csv': b',Country,Surface area (km2),"Population 1,000  (2017)",GDP (M US$),Life Exp Female,Life Exp Male\r\n1,A
```

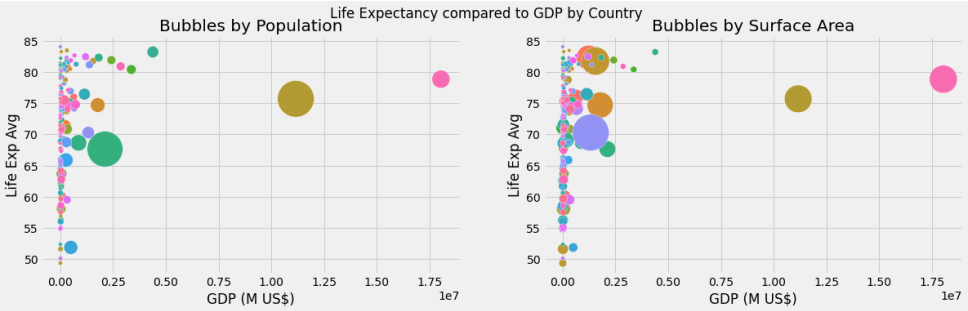
```
cpv = pd.read_csv('country_profile_variables.csv')
cpv = cpv.drop(columns=['Unnamed: 0'])

diff = []
for i in range(len(cpv['Life Exp Female'])):
  diff.append(cpv['Life Exp Female'][i] - cpv['Life Exp Male'][i])

cpv['Life Exp Diff'] = diff

print(cpv.sort_values(['Life Exp Diff'], ascending=False).head(10))

usa_diff = float(cpv['Life Exp Diff'].loc[cpv['Country'] == 'United States of America'])
print('Life Expectancy Female vs Male in USA: {:.2f} years'.format(usa_diff))
```

```
                  Country  Surface area (km2)  ...  Life Exp Male  Life Exp Diff
182  Syrian Arab Republic              185180  ...           64.4           11.9
154    Russian Federation            17098246  ...           64.7           11.2
16                Belarus              207600  ...           66.5           11.2
107             Lithuania               65286  ...           68.5           10.8
102                Latvia               64573  ...           68.8            9.9
195               Ukraine              603500  ...           66.1            9.9
141                 Palau                 459  ...           68.1            9.7
96             Kazakhstan             2724902  ...           64.3            9.6
205              Viet Nam              330967  ...           70.7            9.6
60                Estonia               45227  ...           71.8            9.4

[10 rows x 7 columns]
Life Expectancy Female vs Male in USA: 4.70 years
```

```
average = []
for i in range(len(cpv['Life Exp Female'])):
  average.append((cpv['Life Exp Female'][i] + cpv['Life Exp Male'][i])/2)

cpv['Life Exp Avg'] = average

cpv.head(5)
```

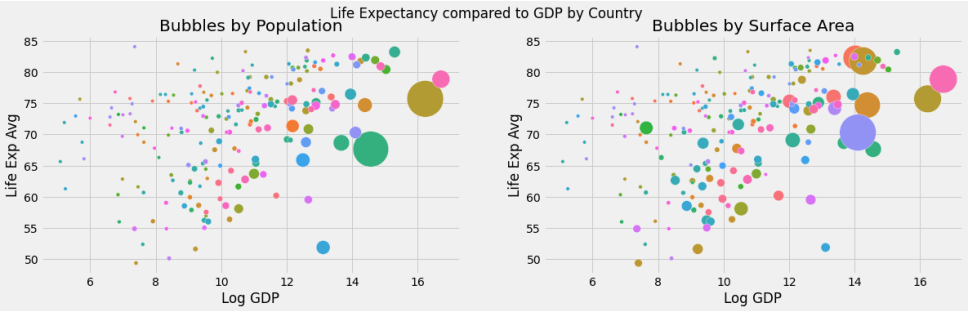
	Country	Surface area (km2)	Population 1,000 (2017)	GDP (M US$)	Life Exp Female	Life Exp Male	Life Exp Diff	Life Exp Avg
0	Afghanistan	652864	35530	20270	63.5	61.0	2.5	62.25
1	Albania	28748	2930	11541	79.9	75.6	4.3	77.75
2	Algeria	2381741	41318	164779	76.5	74.1	2.4	75.30
3	American Samoa	199	56	-99	77.8	71.1	6.7	74.45

```
fig, ax = plt.subplots(1, 2, figsize=(18,5))
fig.suptitle('Life Expectancy compared to GDP by Country')

sns.scatterplot(ax=ax[0], data=cpv, x='GDP (M US$)', y='Life Exp Avg', size='Population 1,000  (2017)', hue='Country', legend=False, sizes=(20, 
ax[0].set_title('Bubbles by Population')
sns.scatterplot(ax=ax[1], data=cpv, x='GDP (M US$)', y='Life Exp Avg', size='Surface area (km2)', hue='Country', legend=False, sizes=(20, 2000))
ax[1].set_title('Bubbles by Surface Area')

plt.show()
```



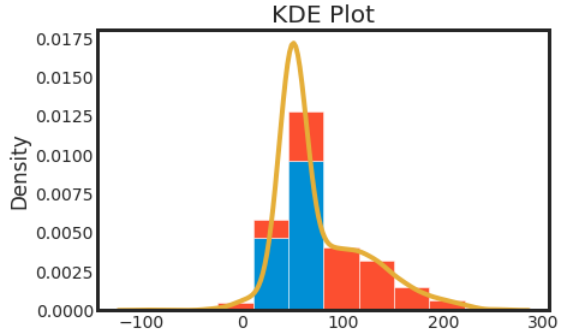

```
cpv['Log GDP'] = np.log(cpv['GDP (M US$)'])

fig, ax = plt.subplots(1, 2, figsize=(18,5))
fig.suptitle('Life Expectancy compared to GDP by Country')

sns.scatterplot(ax=ax[0], data=cpv, x='Log GDP', y='Life Exp Avg', size='Population 1,000  (2017)', hue='Country', legend=False, sizes=(20, 2000
ax[0].set_title('Bubbles by Population')
sns.scatterplot(ax=ax[1], data=cpv, x='Log GDP', y='Life Exp Avg', size='Surface area (km2)', hue='Country', legend=False, sizes=(20, 2000))
ax[1].set_title('Bubbles by Surface Area')

plt.show()
```

```
/usr/local/lib/python3.7/dist-packages/pandas/core/series.py:726: RuntimeWarning: invalid value
  result = getattr(ufunc, method)(*inputs, **kwargs)
```



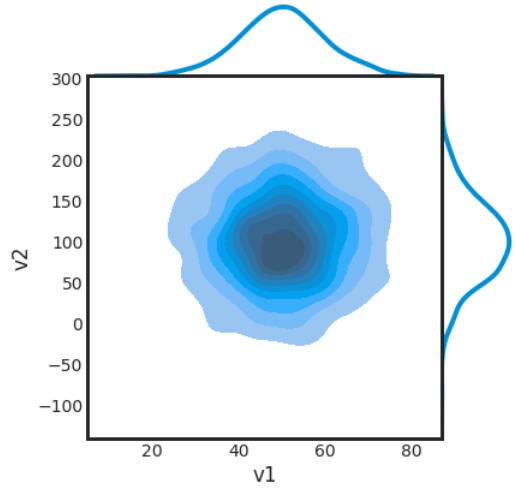

```
##### Problem 2 #####

np.random.seed(1234)

v1 = pd.Series(np.random.normal(50,10,1000), name='v1')
v2 = pd.Series(np.random.normal(100,50,1000), name='v2')

plt.figure()
plt.hist([v1,v2], histtype='barstacked', density=True)
v3 = np.concatenate((v1,v2))
sns.kdeplot(v3)
plt.title('KDE Plot')
```

```
Text(0.5, 1.0, 'KDE Plot')
```




```
sns.set_style('white')
sns.jointplot(v1, v2, kind='kde', space=0, shade=True)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the follow
  FutureWarning
<seaborn.axisgrid.JointGrid at 0x7f183828d210>
```




```
sns.jointplot(v1,v2,kind='hex')
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the follow
  FutureWarning
<seaborn.axisgrid.JointGrid at 0x7f1839263710>
```