ASSIGNMENT - 1

Finding the independent and dependent parameters using Correlation - 1

In [1]:

```
# importing packages pandas and numpy
import pandas as pd
import numpy as np
```

In [2]:

```
# loading the dataset
income = pd.read_csv("Income1.csv")
```

In [3]:

```
income.head()
```

Out[3]:

	Unnamed: 0	Education	Income
0	1	10.000000	26.658839
1	2	10.401338	27.306435
2	3	10.842809	22.132410
3	4	11.244147	21.169841
4	5	11.645485	15.192634

In [4]:

```
# removing column "unnamed: 0" since it is of no use
income = income.drop(['Unnamed: 0'], axis=1)
```

In [5]:

```
income.info()
income.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Education  30 non-null     float64
 1   Income     30 non-null     float64
dtypes: float64(2)
memory usage: 608.0 bytes
```

Out[5]:

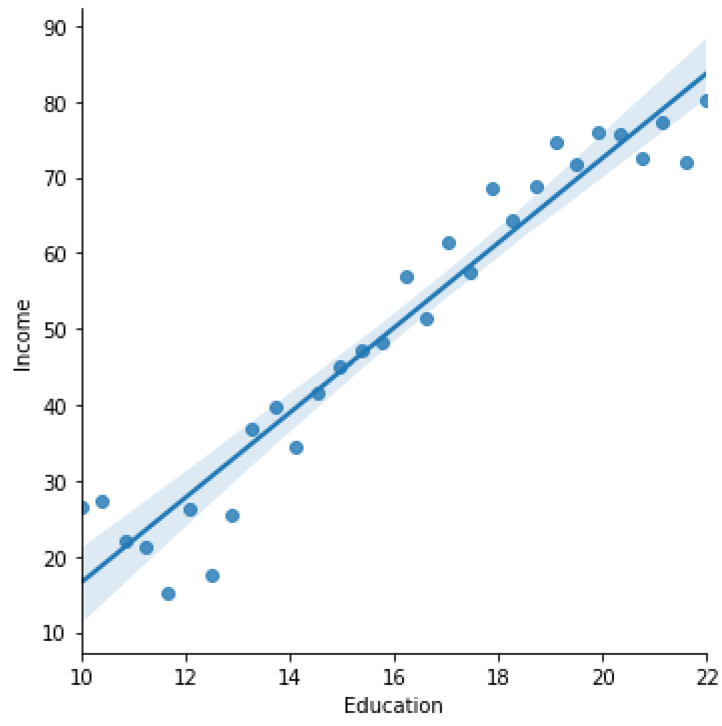
	Education	Income
0	10.000000	26.658839
1	10.401338	27.306435
2	10.842809	22.132410
3	11.244147	21.169841
4	11.645485	15.192634

In [6]:

```
# determining the relation between Education and Income
# importing packages seaborn and matplotlib
import seaborn as sns
import matplotlib.pyplot as plt
sns.lmplot(x="Education", y="Income", data = income)
```

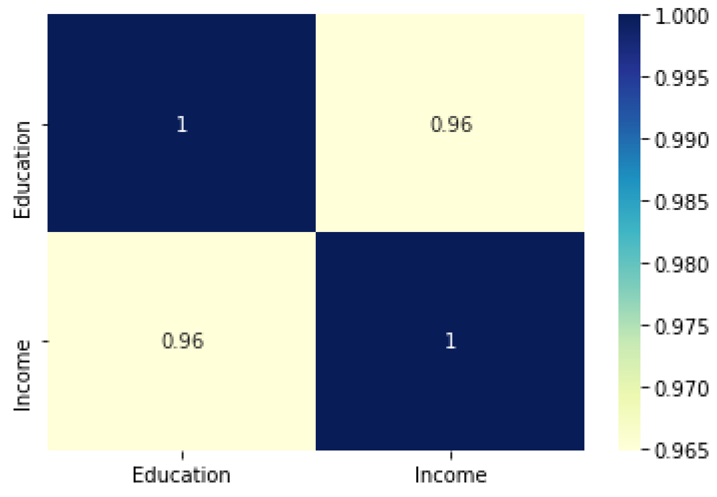
Out[6]:

```
<seaborn.axisgrid.FacetGrid at 0x7fa53919f0d0>
```



In [7]:

```
# visualizing the data using heatmap
sns.heatmap(income.corr(), cmap="YlGnBu", annot = True)
plt.show()
```



As we can see that the education is more correlated to income with correlation value 0.96 Hence Education can be considered as independent parameter in order to predict income which would be the dependent parameter.

In [8]:

```
X = income[['Education']] # taking independent parameter as X
y = income['Income'] # taking dependent parameter as y
```