da-lab-3

March 18, 2023

[ ]:
```
import pandas as pd
import numpy as np
import seaborn as sns
```

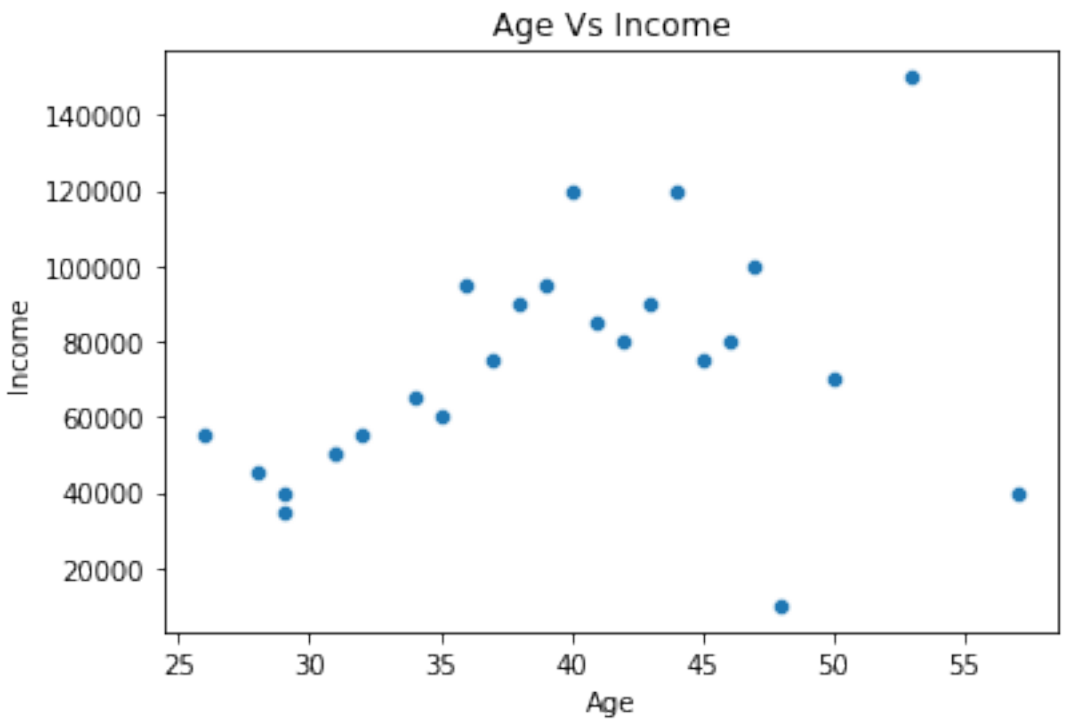
[ ]:
```
df=pd.read_csv('/content/drive/MyDrive/person.csv',index_col=0,na_values=['??
↪','????'])
df.head(5)
```

[ ]:

City	Age	Gender	Marital Status	Income
New York	32	Male	Single	55000
Toronto	45	Female	Married	75000
Paris	28	Male	Single	45000
Berlin	40	Female	Married	120000
London	31	Male	Single	50000

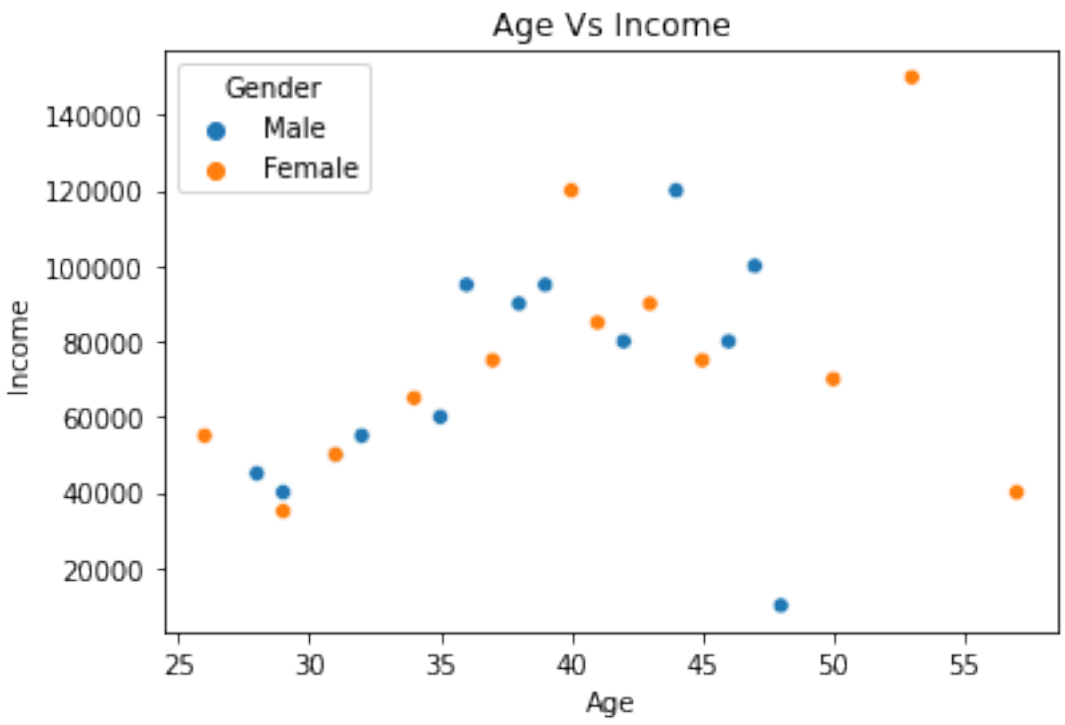
[ ]:
```
#scatter plot- more than 2 variable
ax=sns.scatterplot(data=df,x='Age',y='Income')
ax.set_title('Age Vs Income')

# As we can see in below garph that age of person increasing with Income which␣
↪means that
# as Age of person increseing with the Income of him/her
```

[ ]:
```
Text(0.5, 1.0, 'Age Vs Income')
```



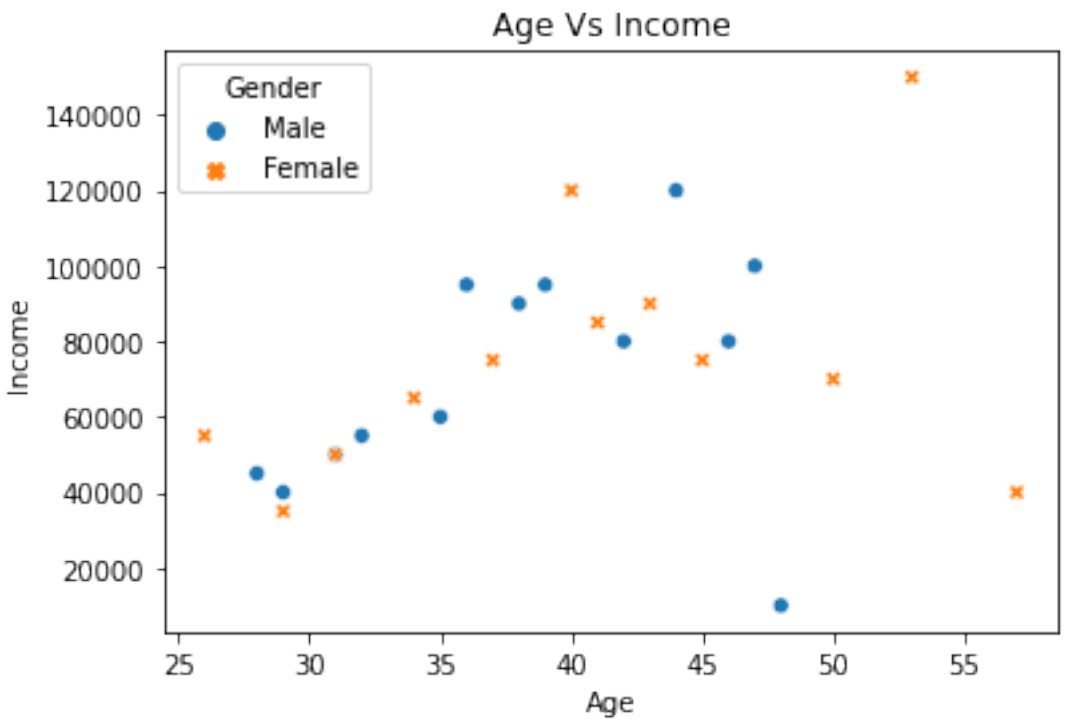
[ ]:
```
#scatter plot- Age Vs Income
ax=sns.scatterplot(data=df,x='Age',y='Income', hue='Gender')
ax.set_title('Age Vs Income')

# As we can see in below garph that Income of Female is increasing with repect␣
 ↪to mans Income
# and the female who is aged are having high income than the other male and␣
 ↪younger female
```

[ ]:
```
Text(0.5, 1.0, 'Age Vs Income')
```



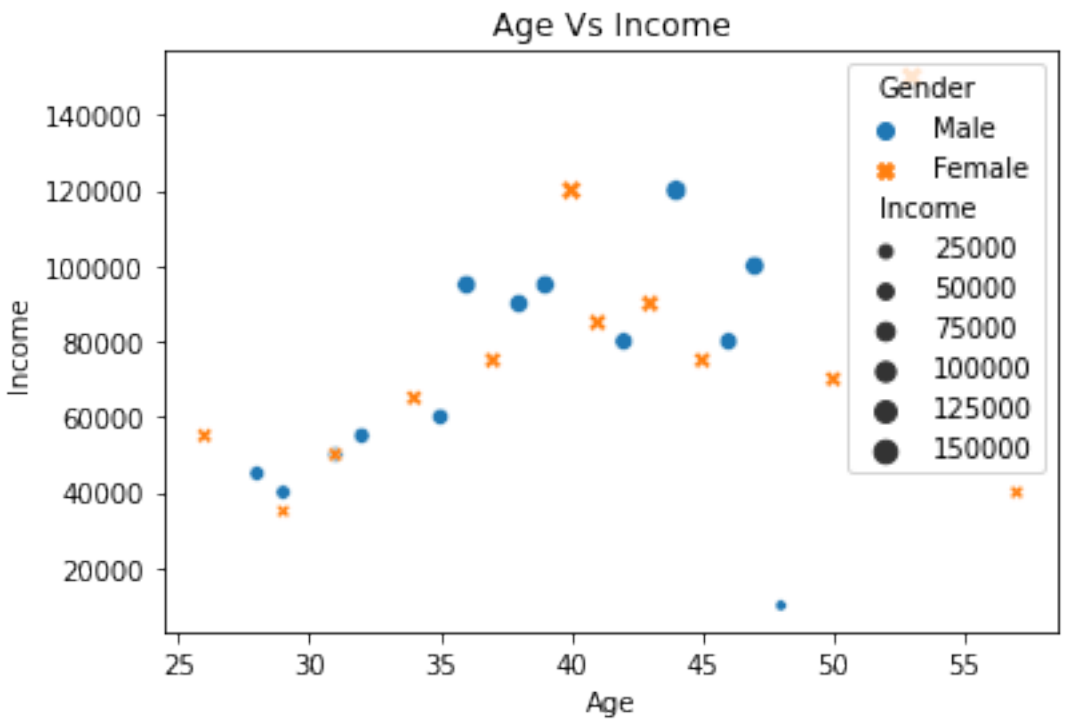
[ ]:
```
#scatter plot- Style parameter
ax=sns.scatterplot(data=df,x='Age',y='Income',hue='Gender',)
ax.set_title('Age Vs Income')

# As we can see in below garph that Income of Female is increasing with repect␣
 ↪to mans Income
# Male is represented by circle and female is respresented by square
```

[ ]:
```
Text(0.5, 1.0, 'Age Vs Income')
```



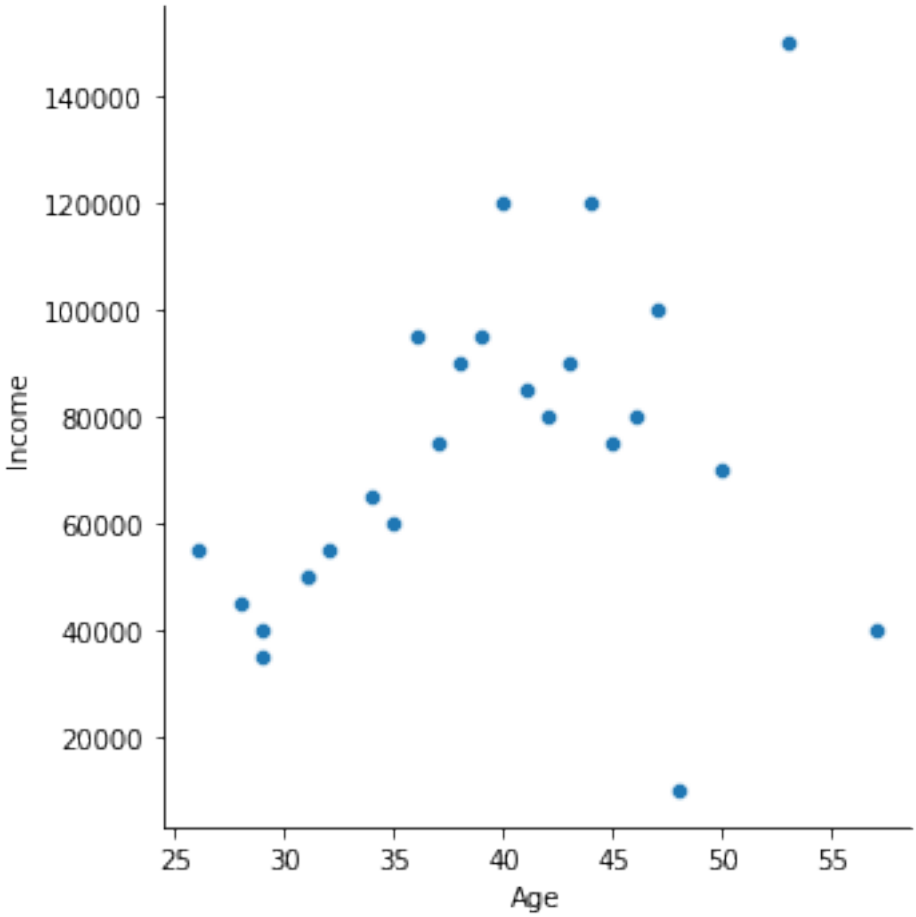
[ ]:
```
#scatter plot- Size parameter
ax=sns.
 ↪scatterplot(data=df,x='Age',y='Income',hue='Gender',style='Gender',size='Income')
ax.set_title('Age Vs Income')

# As we can see in below garph that Income of Female is increasing with repect␣
 ↪to mans Income
# male is respresrented by circle and female is respresented by square
# the size of symbol is representing the range of Income
```

[ ]:
```
Text(0.5, 1.0, 'Age Vs Income')
```



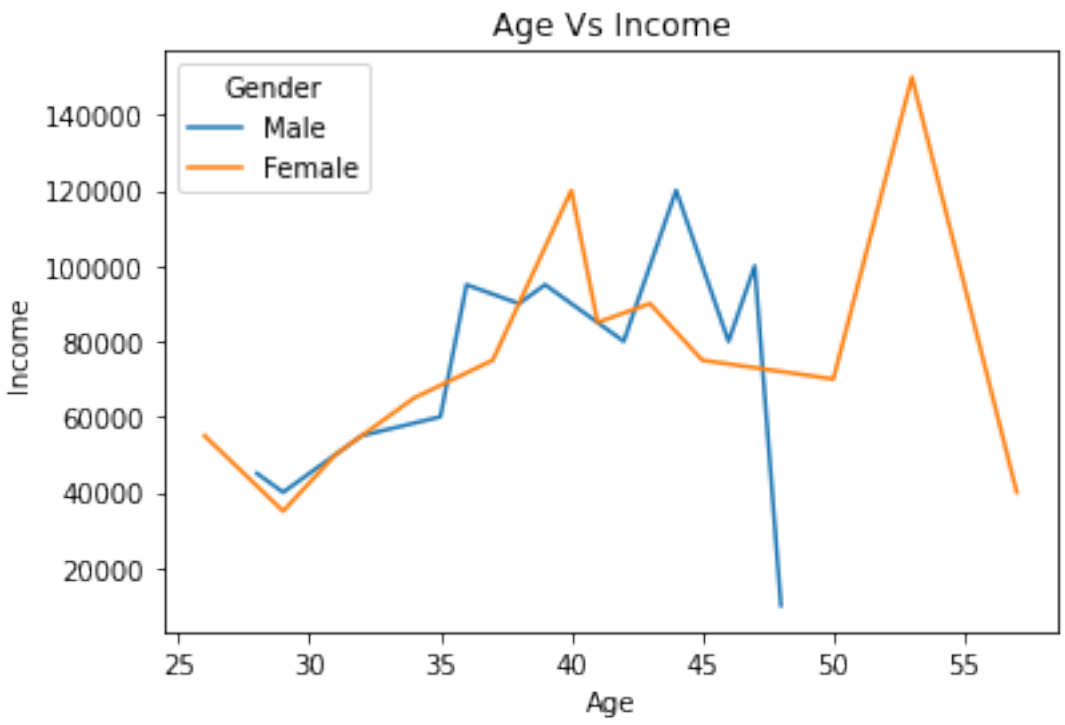
[ ]:
```
#relplot- Age Vs Income
ax=sns.relplot(data=df,x='Age',y='Income')
ax.set_titles('Age Vs Income')

# As we can see in below garph that age of person increasing with Income which␣
 ↪means that
# as Age of person increseeing with the Income of him/her
```

[ ]:
```
<seaborn.axisgrid.FacetGrid at 0x7fe75ca5bfd0>
```



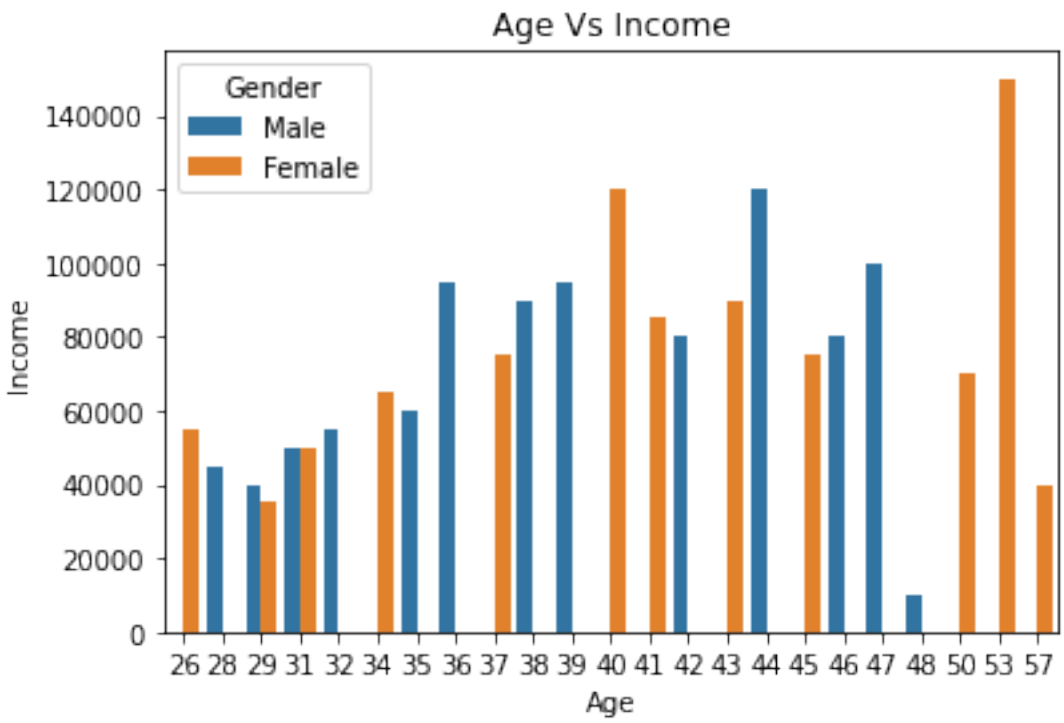
[ ]:
```
#df.reset_index(level=0, inplace=True)
ax= sns.lineplot(data=df,x='Age',y='Income',hue='Gender')
ax.set_title('Age Vs Income')

# In the below graph it is show that aged female income is in growth than of␣
 ↪males income
```

[ ]:
```
Text(0.5, 1.0, 'Age Vs Income')
```



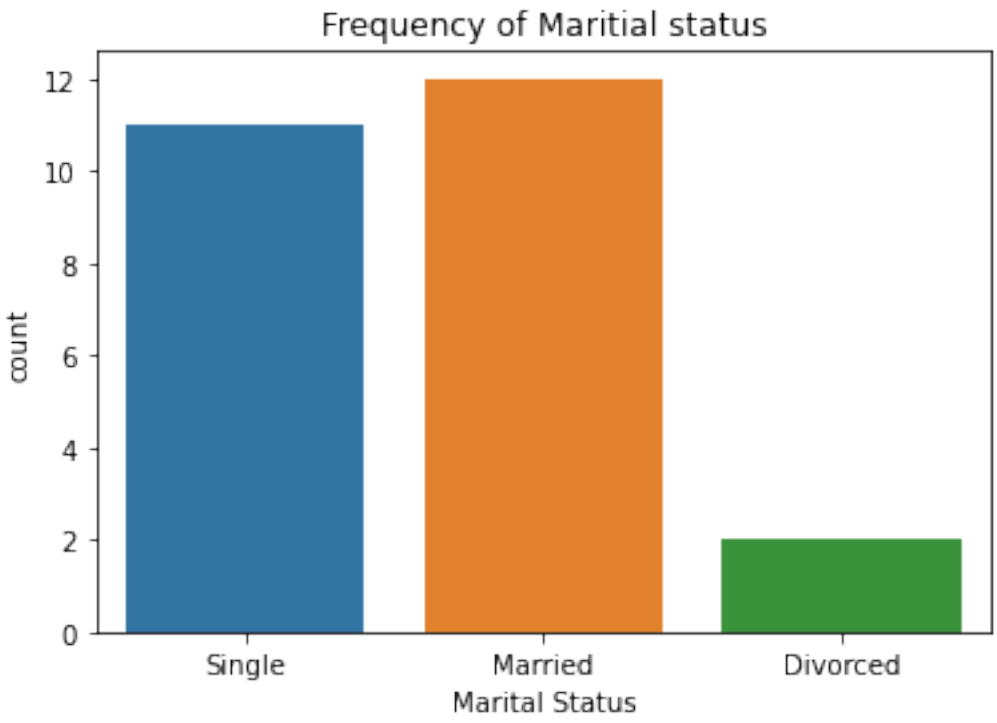
[ ]:
```
ax= sns.barplot(data=df,x='Age',y='Income',hue='Gender')
ax.set_title('Age Vs Income')

# In the below graph the female whose age is 50,53,57 has higher Income than␣
 ↪the other amle and younger female
# males higher income of 110000 Rs
```

[ ]:
```
Text(0.5, 1.0, 'Age Vs Income')
```



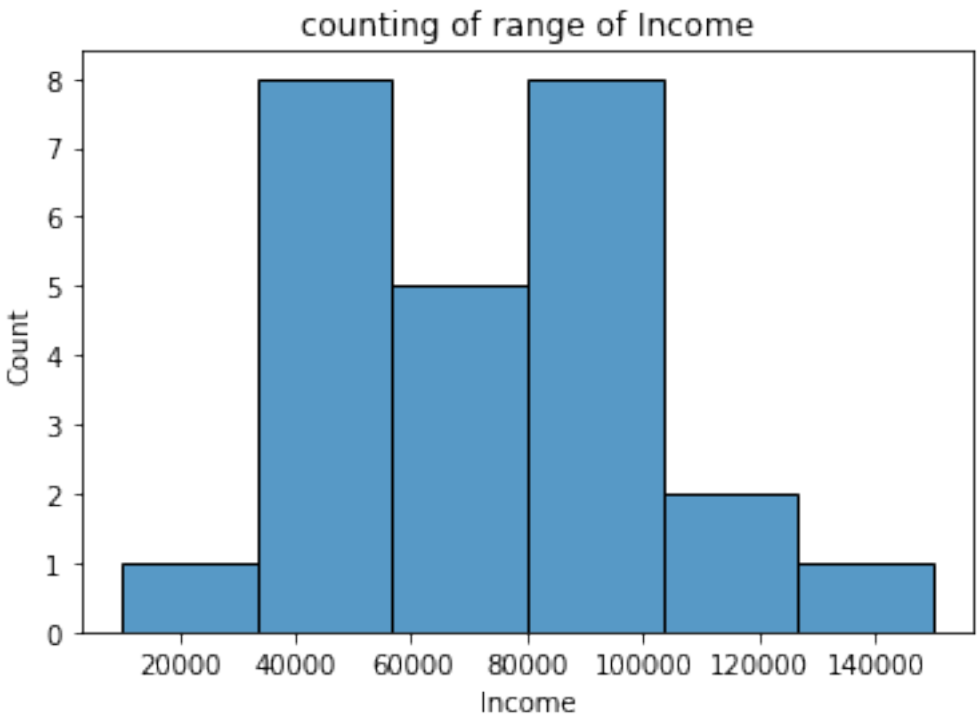
[ ]:
```
ax= sns.countplot(data=df,x='Marital Status')
ax.set_title('Frequency of Maritial status')

# Counting the frequency of maritial and as we can se that married are higher␣
 ↪in counts
# then singles are having less frequency than married and dicorced are the␣
 ↪least among them
```

[ ]:
```
Text(0.5, 1.0, 'Frequency of Maritial status')
```



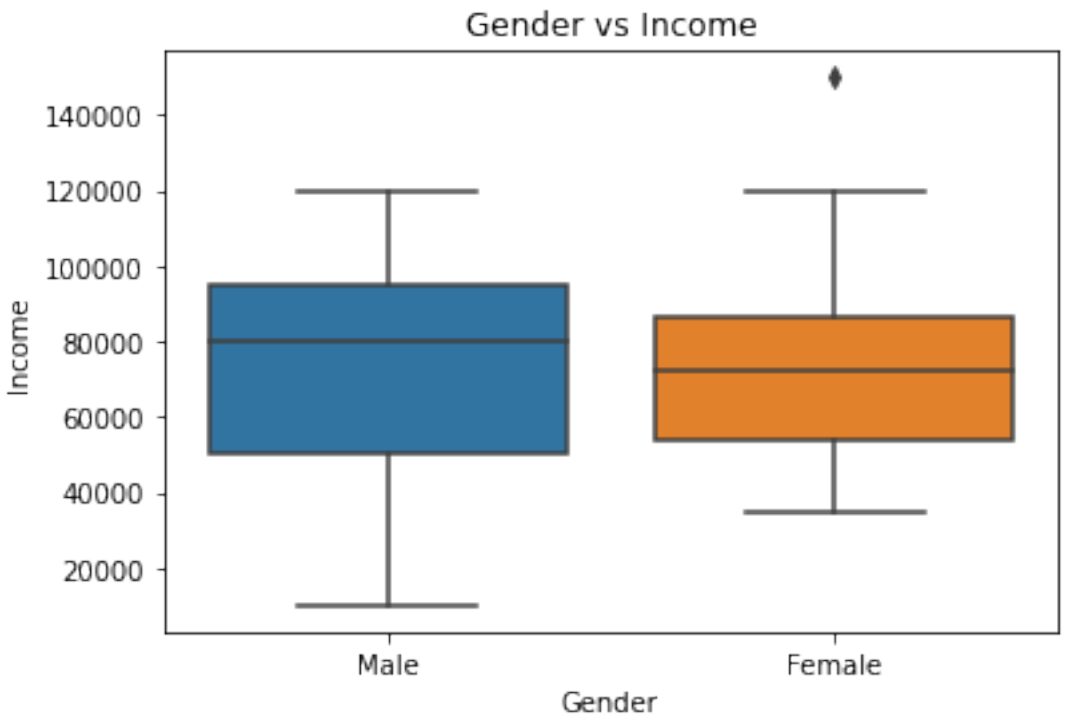
[ ]:
```
#df.reset_index(level=0, inplace=True)
ax= sns.histplot(data=df,x='Income')
ax.set_title('counting of range of Income')

#Counting of range of income and the higher income frequency between␣
 ↪40000-60000 and 80000-100000
```

[ ]:
```
Text(0.5, 1.0, 'counting of range of Income')
```

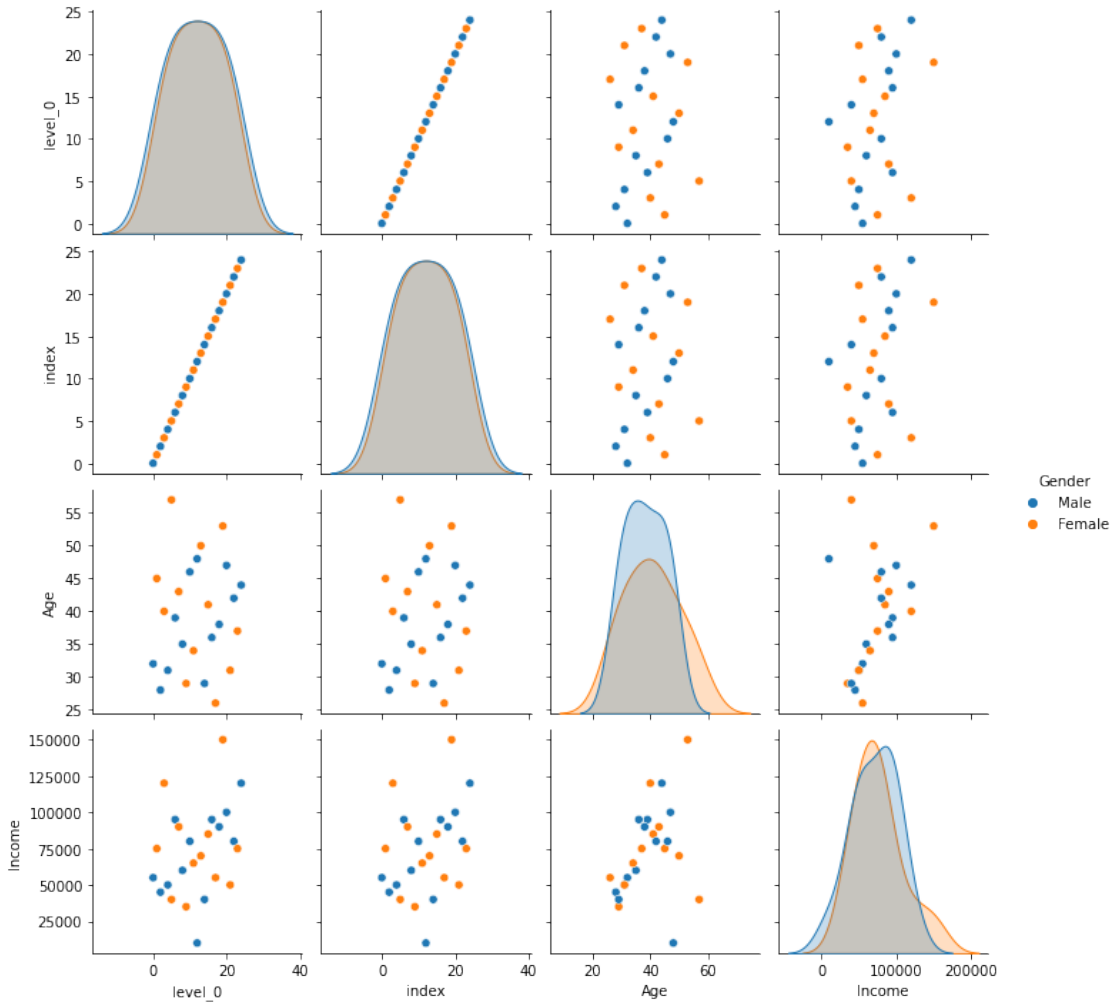


[ ]:
```
ax= sns.boxplot(data=df,y='Income',x='Gender')
ax.set_title('Gender vs Income')

# In the below Graph the medain value of male and female is 80000 and 70000
# most of the male and female income lie in lowest quartile
```

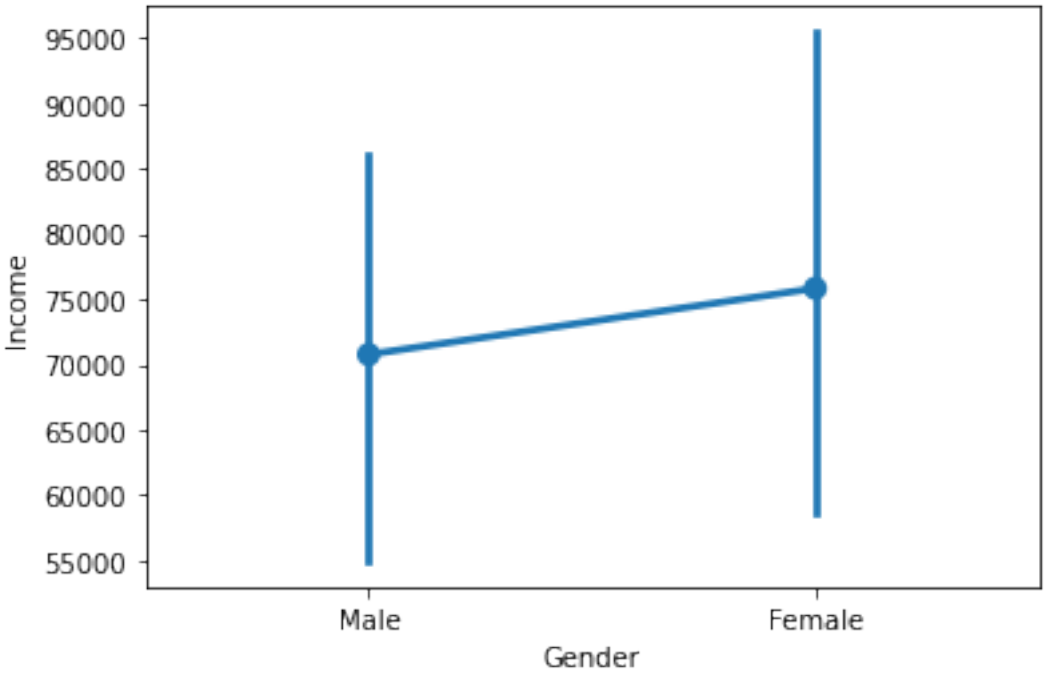
[ ]:
```
Text(0.5, 1.0, 'Gender vs Income')
```



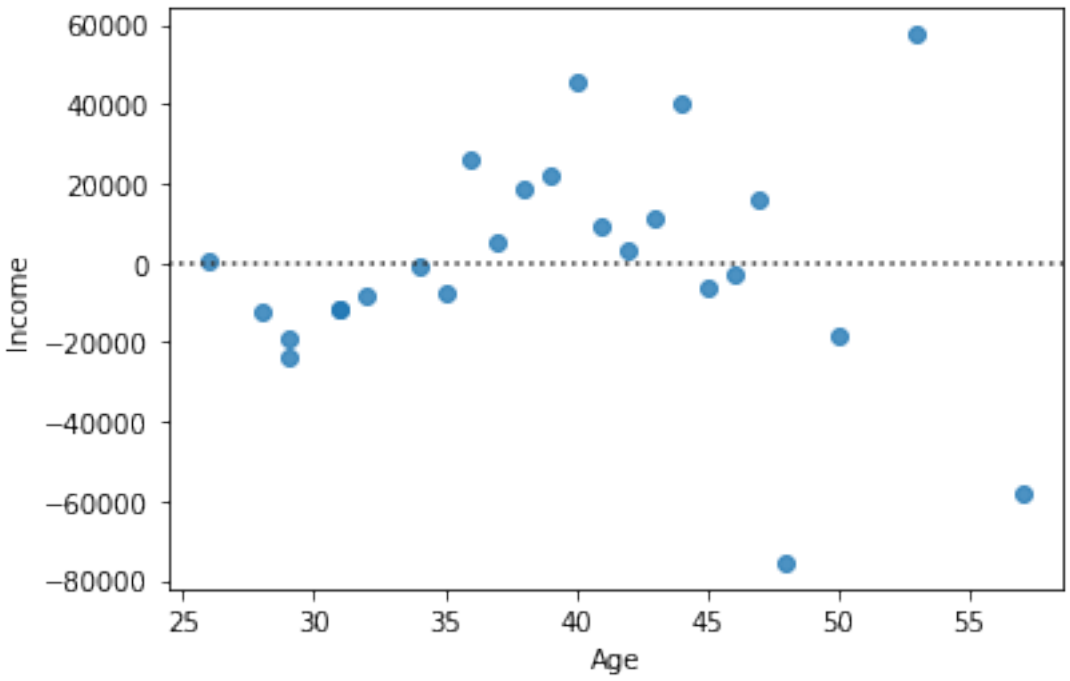
[ ]:
```
ax= sns.pairplot(df,hue='Gender')

# In the graph all the quantitive data is plotted among themself
# and show the relationship among them
```



[ ]:
```
ax= sns.pointplot(data=df,x='Gender',y='Income')

# From the below graph there are relationship among caterigoral(Gender) and␣
 ↪quantative(Income) data
# it is also showing that female income is higher than male it is shown by the␣
 ↪slope of garph
```



[ ]:
```
ax= sns.residplot(data=df,x='Age',y='Income')

# In the below graph the are Incomes value above and below the zero value that␣
 ↪means
# if values are above zero than it overestimate the response variable
# and if values are below Zero than it understiamte the response variable
```


60000
40000
20000
0
−20000
−40000
−60000
−80000
Income
25
30
35
40
45
50
55
Age