

```
In [1]: import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import seaborn as sns
```

load the data from a csv file

```
In [3]: df = pd.read_csv("DoctorVisits - OA.csv")
print(df.head(10))
```

	Unnamed: 0	visits	gender	age	income	illness	reduced	health	private	\
0	1	1	female	0.19	0.55	1	4	1	yes	
1	2	1	female	0.19	0.45	1	2	1	yes	
2	3	1	male	0.19	0.99	3	0	0	no	
3	4	1	male	0.19	0.15	1	0	0	no	
4	5	1	male	0.19	0.45	2	5	1	no	
5	6	1	female	0.19	0.35	5	1	9	no	
6	7	1	female	0.19	0.55	4	0	2	no	
7	8	1	female	0.19	0.15	3	0	6	no	
8	9	1	female	0.19	0.65	2	0	5	yes	
9	10	1	male	0.19	0.15	1	0	0	yes	

	freepoor	freerepat	nchronic	lchronic
0	no	no	no	no
1	no	no	no	no
2	no	no	no	no
3	no	no	no	no
4	no	no	yes	no
5	no	no	yes	no
6	no	no	no	no
7	no	no	no	no
8	no	no	no	no
9	no	no	no	no

```
In [4]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5199 entries, 0 to 5199
Data columns (total 13 columns):
#   Column      Non-Null Count  Dtype
---  --
0   Unnamed: 0    5199 non-null    int64
1   visits       5199 non-null    int64
2   gender       5199 non-null    object
3   age          5199 non-null    float64
4   income       5199 non-null    float64
5   illness      5199 non-null    int64
6   reduced      5199 non-null    int64
7   health       5199 non-null    int64
8   private      5199 non-null    object
9   freepoor     5199 non-null    object
10  freerepat    5199 non-null    object
11  nchronic     5199 non-null    object
12  lchronic     5199 non-null    object
dtypes: float64(2), int64(5), object(6)
memory usage: 527.2+ KB
```

```
In [5]: df.describe()

Out[5]:
```

	Unnamed: 0	visits	age	income	illness	reduced	health
count	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000	5190.000000
mean	2595.500000	0.301734	0.406385	0.583160	1.431985	0.861850	1.217534
std	1490.368279	0.796134	0.204782	0.368907	1.384152	2.887628	2.124266
min	1.000000	0.000000	0.190000	0.000000	0.000000	0.000000	0.000000
25%	1298.250000	0.000000	0.220000	0.250000	0.000000	0.000000	0.000000
50%	2595.500000	0.000000	0.320000	0.550000	1.000000	0.000000	0.000000
75%	3892.750000	0.000000	0.620000	0.900000	2.000000	0.000000	2.000000
max	5190.000000	9.000000	0.720000	1.500000	5.000000	14.000000	12.000000

```
In [6]: df["gender"].value_counts()

Out[6]:
```

female 2792  
male 2488  
Name: gender, dtype: int64

# Find the no:of people based on their count of illeness

```
In [7]: df["illness"].value_counts()

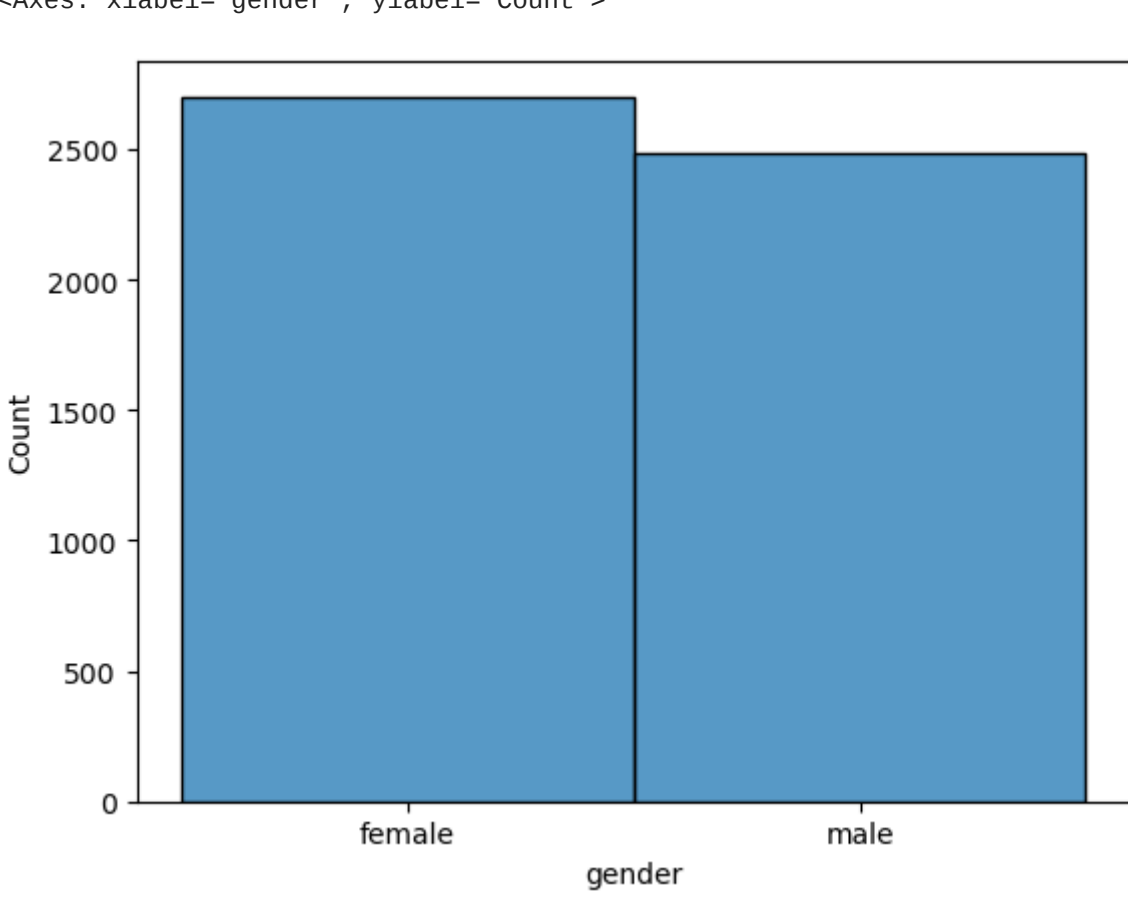
Out[7]:
```

1 1638  
0 1554  
2 946  
3 542  
4 274  
5 236  
Name: illness, dtype: int64

visualize the number of male and female affected by illness

```
In [8]: sns.histplot(df.gender,bins=2)

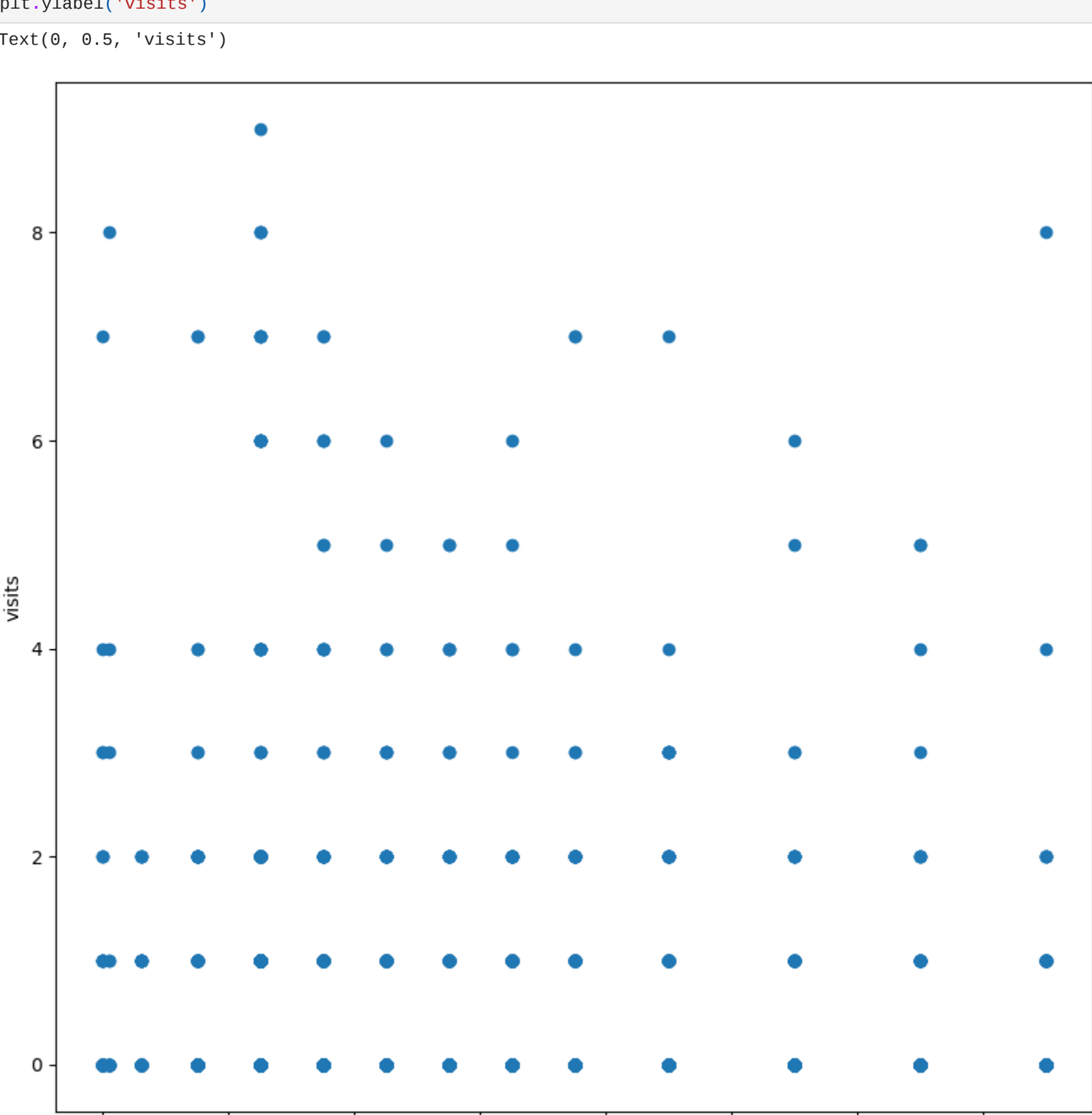
Out[8]: <Axes: xlabel='gender', ylabel='Count'>
```



relation between income and no:of visits to hospital

```
In [9]: plt.figure(figsize=(10,10))
plt.scatter(x='income',y='visits',data=df)
plt.xlabel("income")
plt.ylabel("visits")

Out[9]: Text(0, 0.5, 'visits')
```



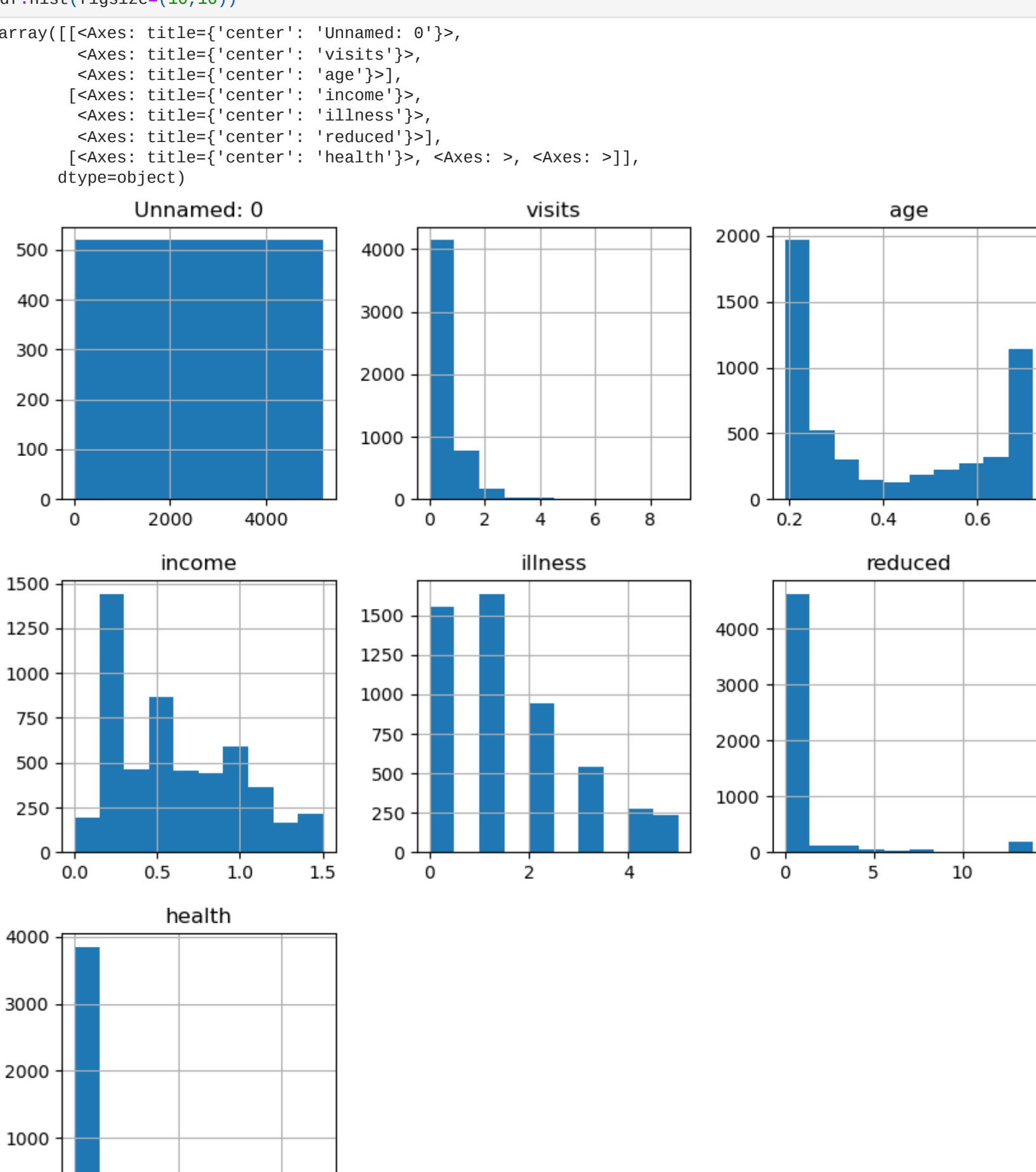
```
In [16]: total_visits=len(df)
total_illnesses=df["illness"].sum()
total_visits=len(df)
total_reduced=df["health"].sum()
total_income=df["income"].sum()
total_health=df["health"].sum()
print('Total illness:',total_illnesses)
print('Total visits:',total_visits)
print('Total reduced:',total_reduced)
print('Total income:',total_income)
print('Total health:',total_health)

Total illness: 7432
Total visits: 5190
Total reduced: 6319
Total income: 738
Total health: 6319
```

visualise of data

```
In [11]: df.hist(figsize=(10,10))

Out[11]: array([[<Axes: title={'center': 'Unnamed: 0'}>,
<Axes: title={'center': 'visits'}>,
<Axes: title={'center': 'age'}>],
[<Axes: title={'center': 'income'}>,
<Axes: title={'center': 'illness'}>,
<Axes: title={'center': 'reduced'}>],
[<Axes: title={'center': 'health'}>, <Axes: >],
dtype=object])
```



```
In [12]: df["income"]=df["income"].astype(int)
averageIncome=df["income"].mean()
print('averageIncome:',averageIncome)

averageIncome: 0.14219653179190753
```

visualise of average income

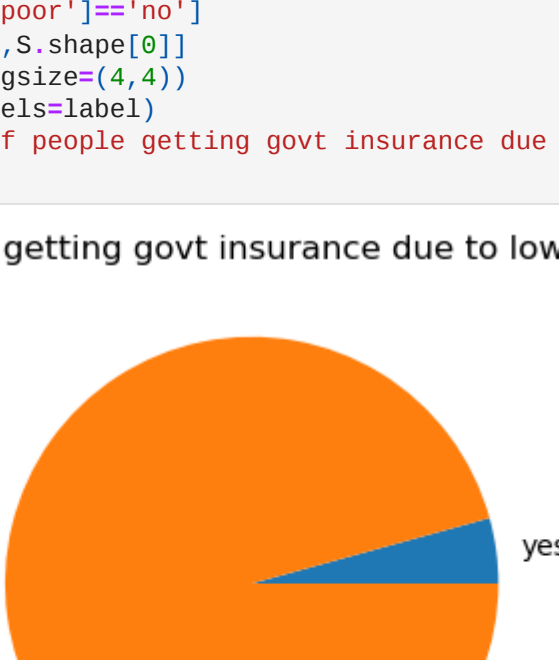
```
In [17]: averageIncome=0.14219653179190753
labels=['averageIncome']
values=[averageIncome]
plt.bar(labels,values,color='yellow')
plt.show()
```



percentage of people getting govt insurance due to low income

```
In [18]: label=['yes','no']
Y=df[df['freepoor']=='yes']
S=df[df['private']=='no']
x=Y.shape[0],S.shape[0]
plt.figure(figsize=(4,4))
plt.pie(x,labels=label)
plt.title('%of people getting govt insurance due to low income')
plt.show()
```

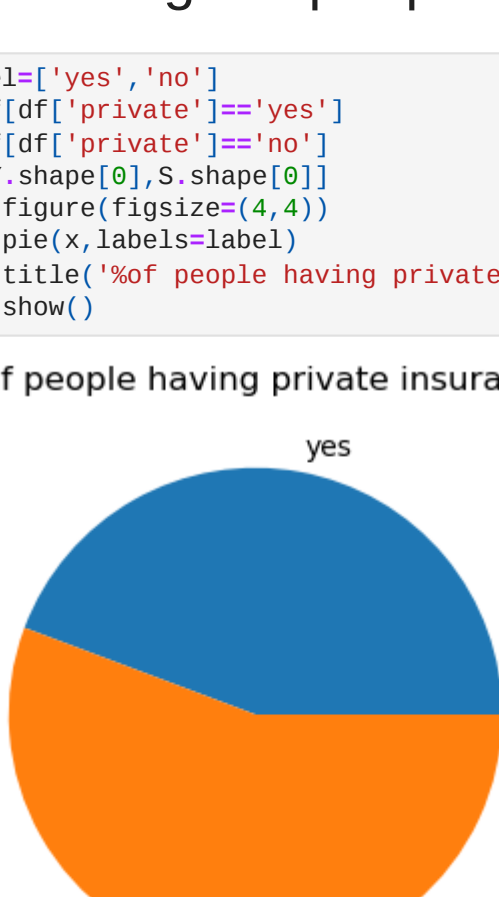
%of people getting govt insurance due to low income



percentage of people having private insurance

```
In [19]: label=['yes','no']
Y=df[df['private']=='yes']
S=df[df['private']=='no']
x=Y.shape[0],S.shape[0]
plt.figure(figsize=(4,4))
plt.pie(x,labels=label)
plt.title('%of people having private insurance')
plt.show()
```

%of people having private insurance



percentage of people getting govt insurance due to old age, disability

```
In [20]: label=['yes','no']
Y=df[df['freerepat']=='yes']
S=df[df['freerepat']=='no']
x=Y.shape[0],S.shape[0]
plt.figure(figsize=(4,4))
plt.pie(x,labels=label)
plt.title('%of people getting govt insurance due to old age, disability')
plt.show()
```

%of people getting govt insurance due to old age, disability

