In [3]:
```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import seaborn as sns
```

Load the dataset and display items

In [4]:
```
df=pd.read_csv("Doctorvisitanalysis.csv") #go to excel paste data and ctrl+s->more options->choosepath(C:\Users
```

Load the dataset and display items

In [5]:
```
df=pd.read_csv("Doctorvisitanalysis.csv")
```

In [6]:
```
print(df.head(15).to_string(index=False))
```

```
 visits gender  age  income  illness  reduced  health private freepoor freerepat nchronic lchronic
      1 female 0.19    0.55        1        4       1     yes       no         no        no        no
      1 female 0.19    0.45        1        2       1     yes       no         no        no        no
      1   male 0.19    0.90        3        0       0      no       no         no        no        no
      1   male 0.19    0.15        1        0       0      no       no         no        no        no
      1   male 0.19    0.45        2        5       1      no       no         no       yes        no
      1 female 0.19    0.35        5        1       9      no       no         no       yes        no
      1 female 0.19    0.55        4        0       2      no       no         no        no        no
      1 female 0.19    0.15        3        0       6      no       no         no        no        no
      1 female 0.19    0.65        2        0       5     yes       no         no        no        no
      1   male 0.19    0.15        1        0       0     yes       no         no        no        no
      1   male 0.19    0.45        1        0       0      no       no         no        no        no
      1   male 0.19    0.25        2        0       2      no       no        yes        no        no
      2   male 0.19    0.55        3       13       1      no       no         no       yes        no
      1   male 0.19    0.45        4        7       6      no       no         no       yes        no
      1   male 0.19    0.25        3        1       0     yes       no         no       yes        no
```

In [11]:
```
print(df.head(15))
```

```
    visits  gender   age  income  illness  reduced  health private freepoor  \
0        1  female  0.19    0.55        1        4       1     yes       no
1        1  female  0.19    0.45        1        2       1     yes       no
2        1    male  0.19    0.90        3        0       0      no       no
3        1    male  0.19    0.15        1        0       0      no       no
4        1    male  0.19    0.45        2        5       1      no       no
5        1  female  0.19    0.35        5        1       9      no       no
6        1  female  0.19    0.55        4        0       2      no       no
7        1  female  0.19    0.15        3        0       6      no       no
8        1  female  0.19    0.65        2        0       5     yes       no
9        1    male  0.19    0.15        1        0       0     yes       no
10       1    male  0.19    0.45        1        0       0      no       no
11       1    male  0.19    0.25        2        0       2      no       no
12       2    male  0.19    0.55        3       13       1      no       no
13       1    male  0.19    0.45        4        7       6      no       no
14       1    male  0.19    0.25        3        1       0     yes       no

   freerepat nchronic lchronic
0         no       no       no
1         no       no       no
2         no       no       no
3         no       no       no
4         no      yes       no
5         no      yes       no
6         no       no       no
7         no       no       no
8         no       no       no
9         no       no       no
10        no       no       no
11       yes       no       no
12        no      yes       no
13        no      yes       no
14        no      yes       no
```

Load the dataset and display items

In [12]:
```
print(df.head(15).to_string(index=False))
```

visits	gender	age	income	illness	reduced	health	private	freepoor	freerepat	nchronic	lchronic
1	female	0.19	0.55	1	4	1	yes	no	no	no	no
1	female	0.19	0.45	1	2	1	yes	no	no	no	no
1	male	0.19	0.90	3	0	0	no	no	no	no	no
1	male	0.19	0.15	1	0	0	no	no	no	no	no
1	male	0.19	0.45	2	5	1	no	no	no	yes	no
1	female	0.19	0.35	5	1	9	no	no	no	yes	no
1	female	0.19	0.55	4	0	2	no	no	no	no	no
1	female	0.19	0.15	3	0	6	no	no	no	no	no
1	female	0.19	0.65	2	0	5	yes	no	no	no	no
1	male	0.19	0.15	1	0	0	yes	no	no	no	no
1	male	0.19	0.45	1	0	0	no	no	no	no	no
1	male	0.19	0.25	2	0	2	no	no	yes	no	no
2	male	0.19	0.55	3	13	1	no	no	no	yes	no
1	male	0.19	0.45	4	7	6	no	no	no	yes	no
1	male	0.19	0.25	3	1	0	yes	no	no	yes	no

2. Display complete info of dataset

In [13]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5190 entries, 0 to 5189
Data columns (total 12 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   visits     5190 non-null   int64  
 1   gender     5190 non-null   object 
 2   age        5190 non-null   float64
 3   income     5190 non-null   float64
 4   illness    5190 non-null   int64  
 5   reduced    5190 non-null   int64  
 6   health     5190 non-null   int64  
 7   private    5190 non-null   object 
 8   freepoor   5190 non-null   object 
 9   freerepat  5190 non-null   object 
 10  nchronic   5190 non-null   object 
 11  lchronic   5190 non-null   object 
dtypes: float64(2), int64(4), object(6)
memory usage: 486.7+ KB
```

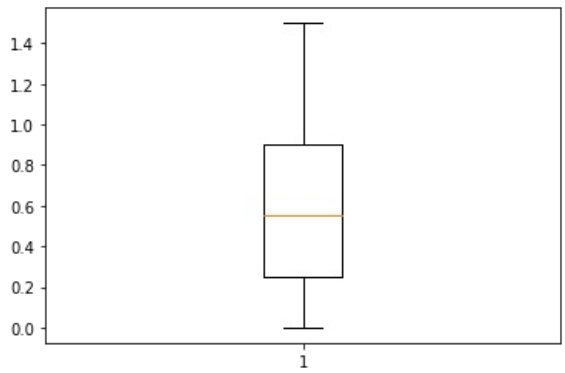
3.Find out total no of people based on theit illness

In [14]:
```
df['illness'].value_counts()
```

Out[14]:
```
1    1638
0    1554
2     946
3     542
4     274
5     236
Name: illness, dtype: int64
```

Visualize and analyse the max,min,medium income

In [15]:
```
y=list(df.income)
plt.boxplot(y)
plt.show()
```



In [17]:
```
df['gender'].value_counts()
```

Out[17]:
```
female    2702
male      2488
Name: gender, dtype: int64
```

Find out the no of days of reduced activity of male and femal

Find out the no of days of reduced activity of male and femal separately due to illness

In [18]:
```
df.groupby(['gender','reduced']).mean()
```

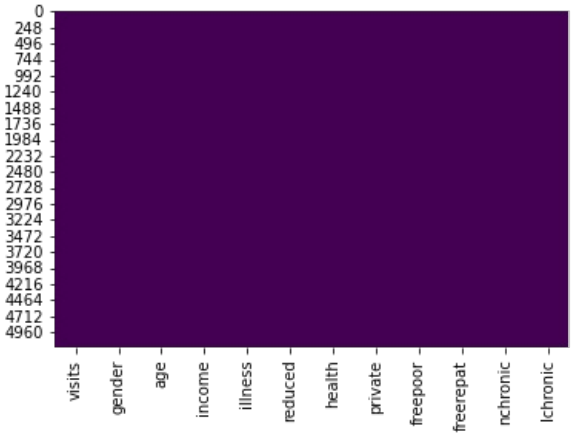
Out[18]:

gender	reduced	visits	age	income	illness	health
female	0	0.229322	0.465755	0.482735	1.462144	1.115098
	1	0.400000	0.325684	0.542105	2.242105	1.610526
	2	0.672727	0.391455	0.560182	2.236364	1.781818
	3	1.333333	0.403111	0.516000	2.733333	1.733333
	4	0.851852	0.458889	0.466667	2.222222	2.074074
	5	1.444444	0.401667	0.614444	2.222222	2.500000
	6	1.363636	0.426364	0.622727	2.363636	1.363636
	7	1.384615	0.436154	0.473462	2.653846	2.230769
	8	1.090909	0.471818	0.404545	2.181818	4.000000
	9	0.500000	0.570000	0.825000	3.000000	1.000000
	10	2.142857	0.512857	0.421429	2.571429	2.000000
	12	2.000000	0.720000	0.250000	3.500000	5.500000
	13	4.000000	0.720000	0.300000	4.500000	3.500000
	14	1.543103	0.551724	0.427586	2.534483	4.112069
male	0	0.136007	0.344703	0.694398	1.099585	0.924850
	1	0.304878	0.286220	0.676341	1.743902	1.256098
	2	0.471698	0.343585	0.653019	2.358491	1.547170
	3	0.724138	0.334138	0.741379	2.137931	1.689655
	4	0.722222	0.309444	0.869444	2.055556	2.000000
	5	1.136364	0.331818	0.570455	2.272727	2.818182
	6	0.833333	0.340000	0.591667	2.500000	2.000000
	7	0.750000	0.314167	0.655000	2.583333	4.333333
	8	1.333333	0.365000	0.833333	2.666667	2.000000
	9	2.200000	0.310000	0.392000	2.400000	2.000000
	10	1.800000	0.480000	0.590000	2.600000	4.600000
	11	5.000000	0.320000	1.000000	1.500000	0.500000
	12	2.000000	0.370000	0.515000	1.500000	1.000000
	13	4.000000	0.510000	0.350000	3.333333	2.333333
	14	1.555556	0.476806	0.598611	2.375000	3.527778

visualize is there any missing values based on heatmaps

In [19]:
```
sns.heatmap(df.isnull(),cbar=False,cmap='viridis')
```

Out[19]:
```
<AxesSubplot:>
```



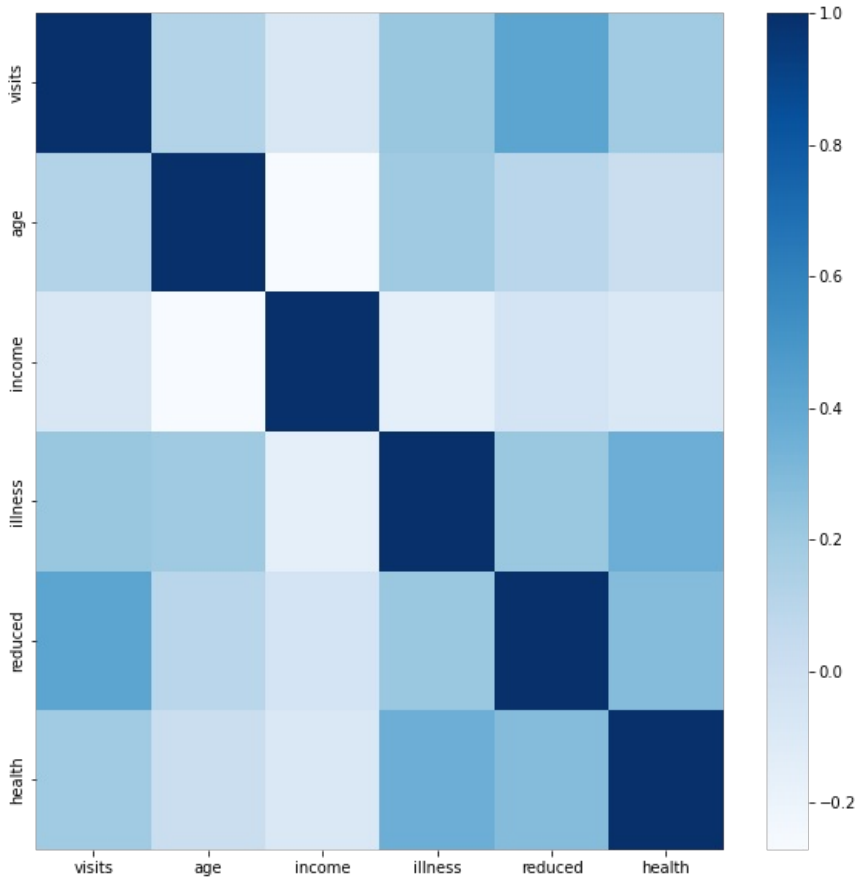
In [38]:
```
df.isnull().sum()
```

Out[38]:
```
visits       0
gender       0
age          0
income       0
illness      0
reduced      0
health       0
private      0
freepoor     0
freerepat    0
nchronic     0
lchronic     0
dtype: int64
```

Find out the correlation between variables in the given dataset correlation between different variables

In [29]:
```
plt.figure(figsize=(10,10))
sns.heatmap(df.corr(),cbar=True,cmap='Blues')
```

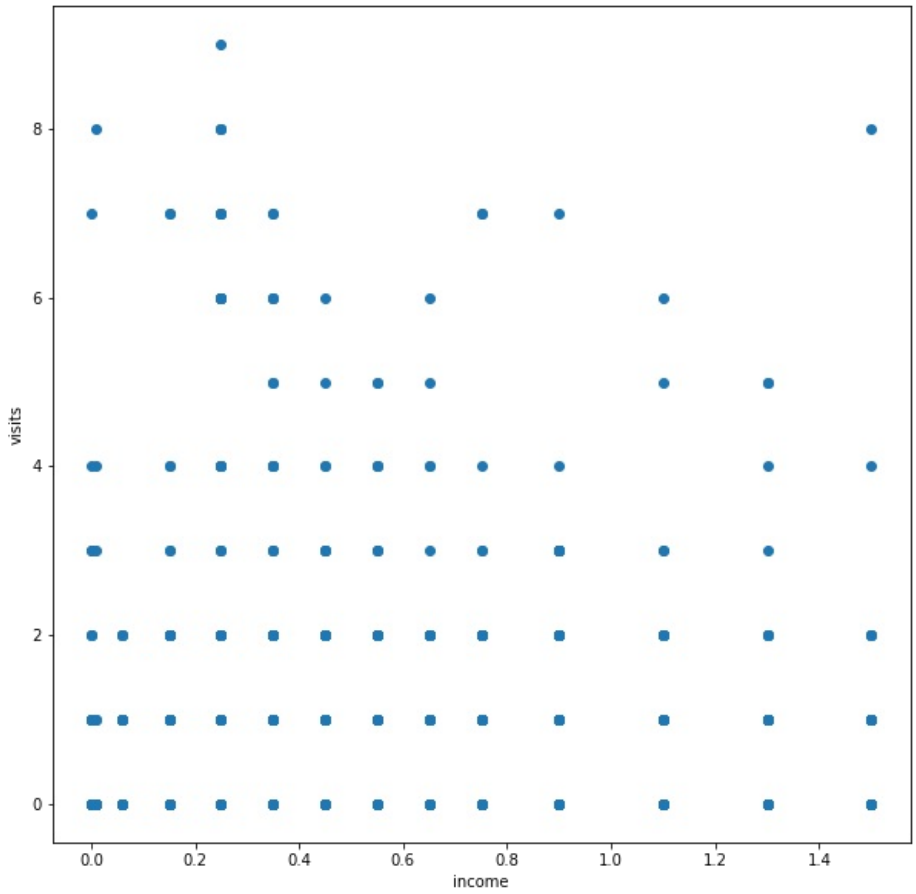
Out[29]:
```
<AxesSubplot:>
```



Analyse how the income of a patient affects the no of visits to the hospital

In [7]:
```
plt.figure(figsize=(10,10))
plt.scatter(x='income',y='visits',data=df)
plt.xlabel('income')
plt.ylabel('visits')
```

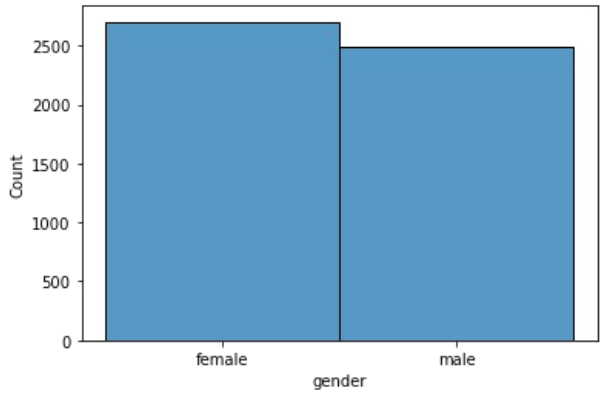
Out[7]:
```
Text(0, 0.5, 'visits')
```



Count and visualizze number of males and females affected by lines

In [32]:
```
sns.histplot(df.gender,bins=2)
```

Out[32]:
```
<AxesSubplot:xlabel='gender', ylabel='Count'>
```



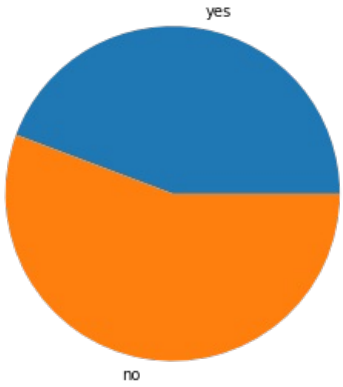
Visualize the percentage of people getting govt health insurance,due to low income,due to oldage,and also due to private health insurance

In [8]:
```
label=['yes','no']
```

```
Y=df[df['freepoor']=='yes']
N=df[df['freepoor']=='no']
x=[Y.shape[0],N.shape[0]]
plt.figure(figsize=(5,5))
plt.pie(x,labels=label)
plt.title("% of people getting govt health insurance due to low income")
plt.show()
label=['yes','no']
Y=df[df['private']=='yes']
N=df[df['private']=='no']
x=[Y.shape[0],N.shape[0]]
plt.figure(figsize=(5,5))
plt.pie(x,labels=label)
plt.title("% of people getting private health insurance due to low income")
plt.show()
label=['yes','no']
Y=df[df['freerepat']=='yes']
N=df[df['freerepat']=='no']
x=[Y.shape[0],N.shape[0]]
plt.figure(figsize=(5,5))
plt.pie(x,labels=label)
plt.title("% of people getting private health insurance due to oldage,disability or vetran status")
plt.show()
```

% of people getting govt health insurance due to low income

% of people getting private health insurance due to low income

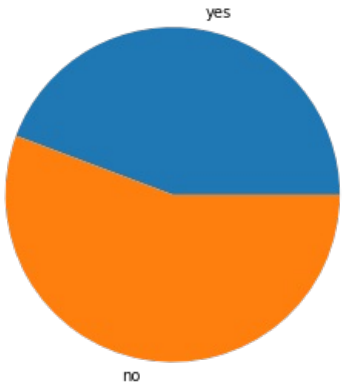


% of people getting private health insurance due to oldage,disability or vetran status

In [34]:
```
label=['yes','no']
Y=df[df['private']=='yes']
N=df[df['private']=='no']
x=[Y.shape[0],N.shape[0]]
plt.figure(figsize=(5,5))
plt.pie(x,labels=label)
```

```
plt.title("% of people getting private health insurance due to low income")
plt.show()
```

% of people getting private health insurance due to low income

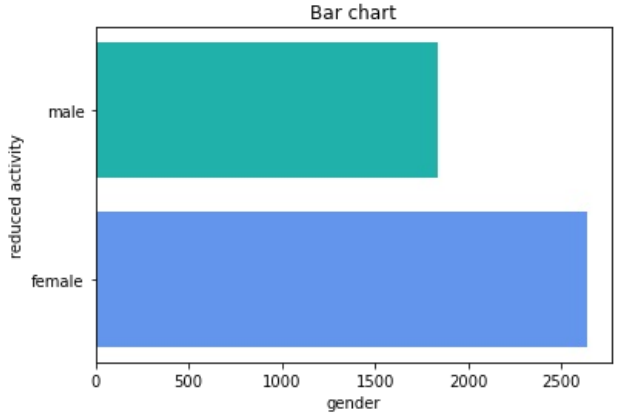


In [35]:
```
label=['yes','no']
Y=df[df['freerepat']=='yes']
N=df[df['freerepat']=='no']
x=[Y.shape[0],N.shape[0]]
plt.figure(figsize=(5,5))
plt.pie(x,labels=label)
plt.title("% of people getting private health insurance due to oldage,disability or vetran status")
plt.show()
```

% of people getting private health insurance due to oldage,disability or vetran status

Plot a horizontal bar chart to analyze the reduced days of activity due to illness based on gender

In [37]:
```
db=df.groupby('gender')['reduced'].sum().to_frame().reset_index()
plt.barh(db['gender'],db['reduced'],color=['cornflowerblue','lightseagreen'])
plt.title('Bar chart')
plt.xlabel('gender')
plt.ylabel('reduced activity')
plt.show()
```



In [ ]:
```

```