```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt


haberman = pd.read_csv("C:\\Users\\SaifzWorld.Lenovo-PC\\Downloads\\haberman.csv")
```

```python
haberman
```

	age	year	nodes	status
0	30	64	1	1
1	30	62	3	1
2	30	65	0	1
3	31	59	2	1
4	31	65	4	1
...	...	...	...	...
301	75	62	1	1
302	76	67	0	1
303	77	65	3	1
304	78	65	1	2
305	83	58	2	2

306 rows × 4 columns

haberman.plot(kind='scatter', x='age', y='status') ; plt.show()

```python
haberman.describe()
#age means age of the person at the time of operation
#here year means year of operation 1900
#nodes means how many positive nodes are detected
#status 1 means patient survive 5 and more than five ,2 mean less than 5
```

	age	year	nodes	status
count	306.000000	306.000000	306.000000	306.000000
mean	52.457516	62.852941	4.026144	1.264706
std	10.803452	3.249405	7.189654	0.441899
min	30.000000	58.000000	0.000000	1.000000
25%	44.000000	60.000000	0.000000	1.000000
50%	52.000000	63.000000	1.000000	1.000000
75%	60.750000	65.750000	4.000000	2.000000
max	83.000000	69.000000	52.000000	2.000000

```
haberman.iloc[:,:3].plot(kind="hist",bins=10,subplots=True,legend=True)
```

Out[37]:

```
array([<AxesSubplot:ylabel='Frequency'>, <AxesSubplot:ylabel='Frequency'>,
       <AxesSubplot:ylabel='Frequency'>], dtype=object)
```
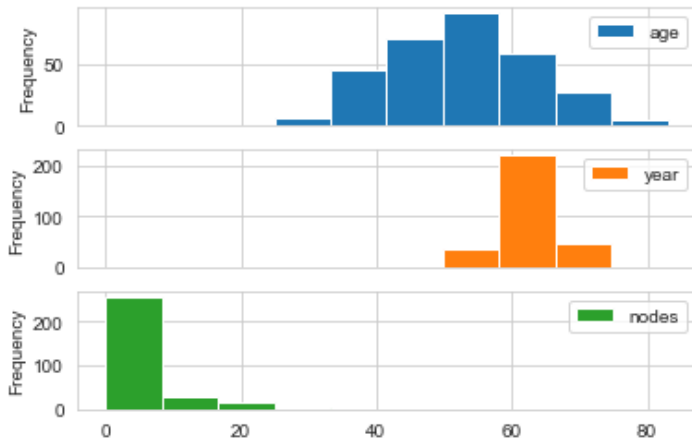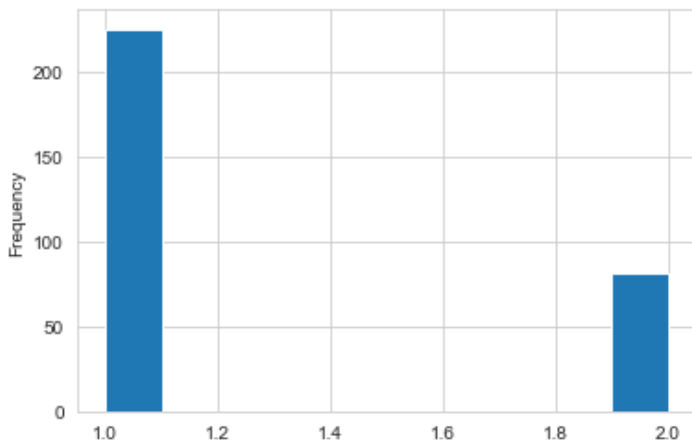

In [38]:

```
haberman["status"].plot(kind="hist")
#more patient had lived there life more than 5 after  operation
```

Out[38]:
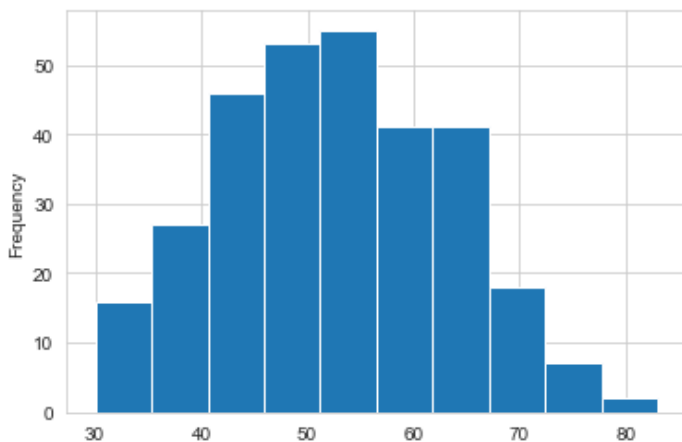
```
<AxesSubplot:ylabel='Frequency'>
```


In [39]:

```
haberman["age"].plot(kind="hist")
#frequency of having operation in age 45-55 is high
```

Out[39]:

```
<AxesSubplot:ylabel='Frequency'>
```


In [40]:

```
haberman["nodes"].plot(kind="hist")
# no of nodes ,this may have outliers
```

Out[40]:

```
<AxesSubplot:ylabel='Frequency'>
```
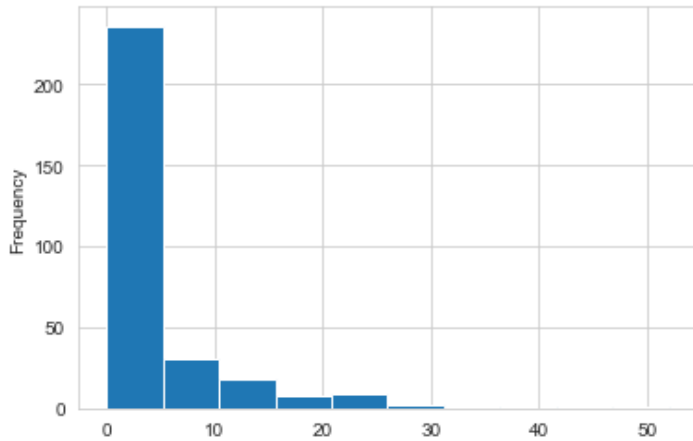

In [42]:

```
haberman["year"].plot(kind="hist")
# in 1958 to 1959 more operation has taken place
```
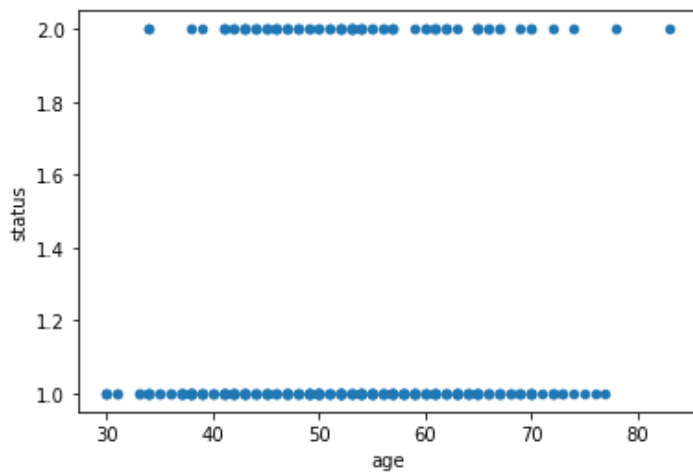
Out[42]:
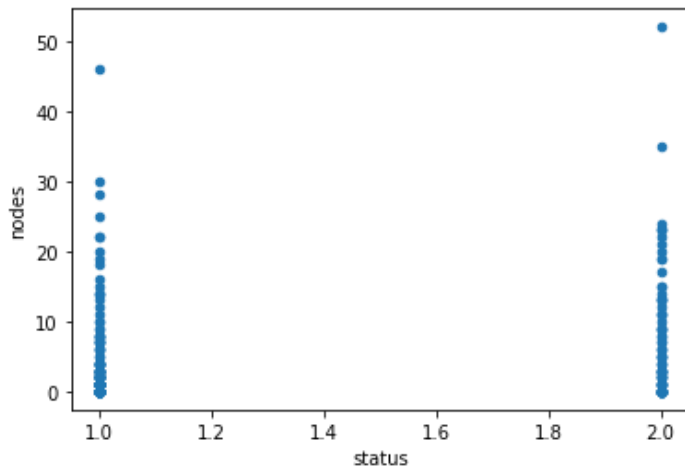
```
<AxesSubplot:ylabel='Frequency'>
```
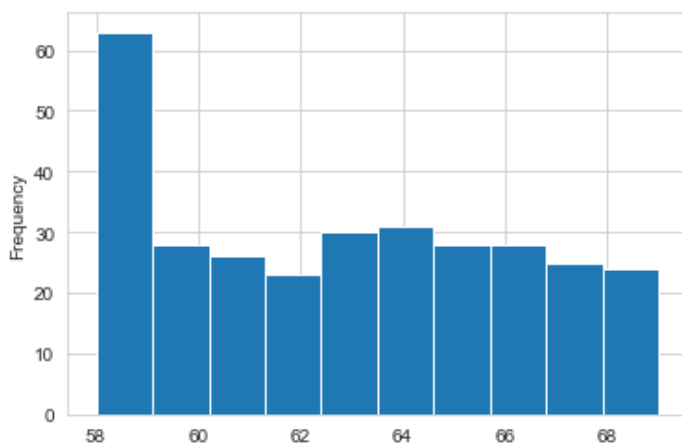

In [24]:

```
haberman.plot(kind='scatter', x='age', y='status') ;
plt.show()
#age and status doesnt show effective info
```
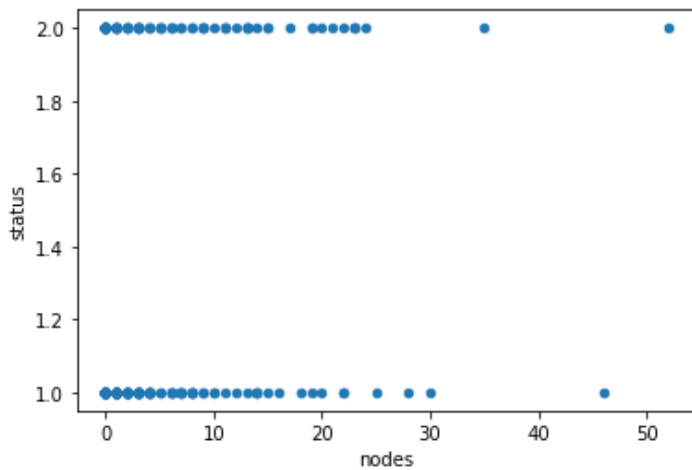

In [27]:

```
haberman.plot(kind='scatter', x='status', y='nodes') ;
```

```
plt.show()
#same with nodes and status nothing informative
```


In [28]:

```
haberman.plot(kind='scatter', x='nodes', y='status') ;
plt.show()
#same with nodes and status nothing informative
```


In [30]:

```
sns.set_style("whitegrid");
sns.FacetGrid(haberman, hue="status", size=4) \
    .map(plt.scatter, "age", "nodes") \
    .add_legend();
plt.show();
#using this two we can distinguish status
```
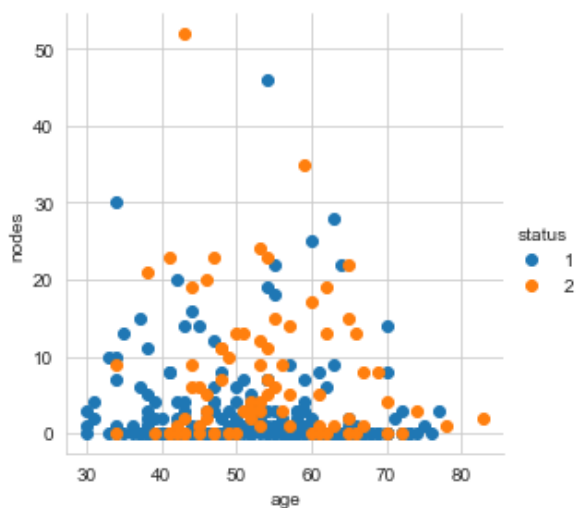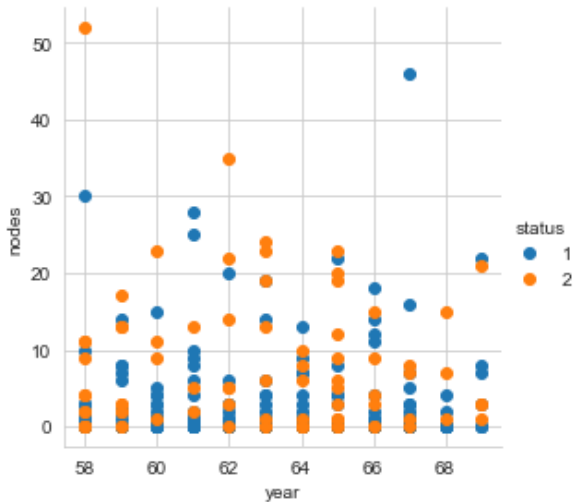
```python
sns.set_style("whitegrid");
sns.FacetGrid(haberman, hue="status", size=4) \
   .map(plt.scatter, "year", "nodes") \
   .add_legend();
plt.show();
#not informative
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\axisgrid.py:316: UserWarning: The `siz
e` parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
```
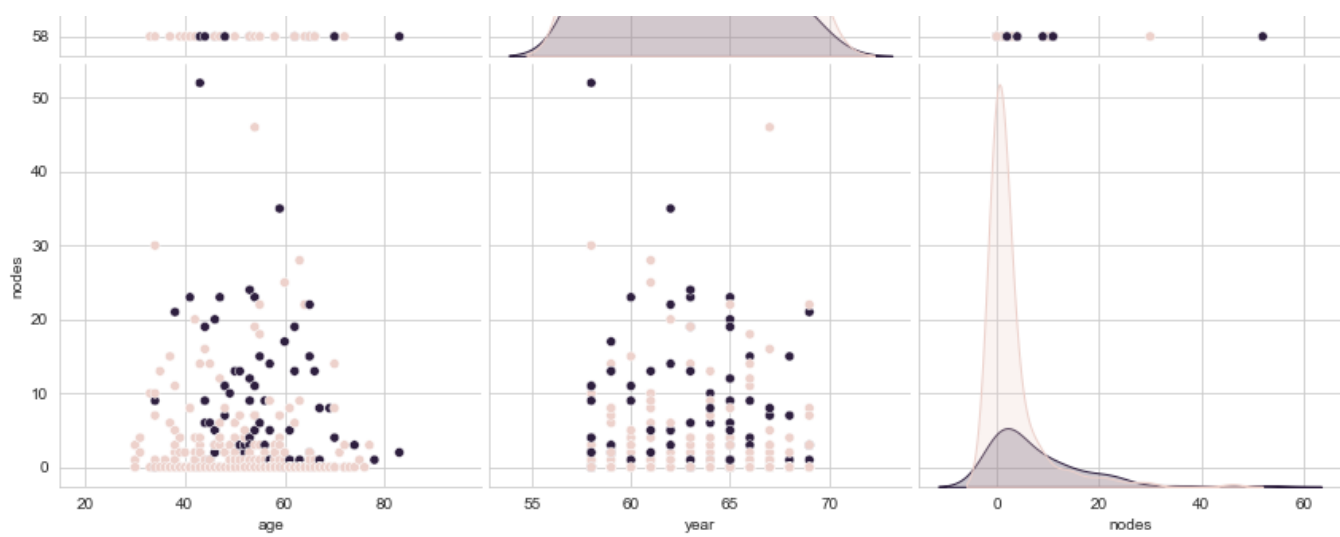
```python
sns.pairplot(haberman,hue="status",size=4)
#&&& by looking node and age we can conclude a litle that age <= 40 and node <= 17 than s
tatus 1
#and if node > 30 and age > 80  status 2 &&&
```
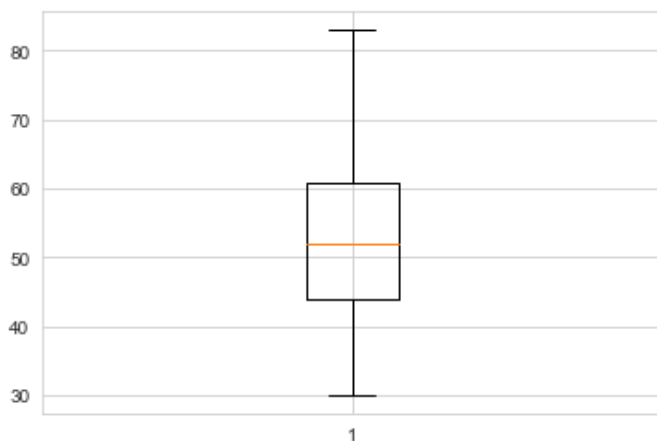
```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\axisgrid.py:1912: UserWarning: The `si
ze` parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
```

Out[33]:

```
<seaborn.axisgrid.PairGrid at 0x8ff7627970>
```

```
plt.boxplot("age",data=haberman)
#the 25% to 75% age lies between 43 to 62
```

```
{'whiskers': [<matplotlib.lines.Line2D at 0x8ffaa420a0>,
  <matplotlib.lines.Line2D at 0x8ffaa42400>],
 'caps': [<matplotlib.lines.Line2D at 0x8ffaa42760>,
  <matplotlib.lines.Line2D at 0x8ffaa42ac0>],
 'boxes': [<matplotlib.lines.Line2D at 0x8ffaa35d00>],
 'medians': [<matplotlib.lines.Line2D at 0x8ffaa42e20>],
 'fliers': [<matplotlib.lines.Line2D at 0x8ffaa4c1c0>],
 'means': []}
```
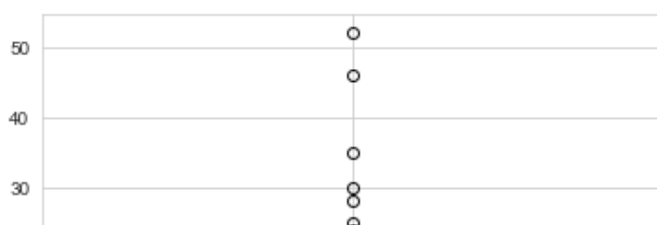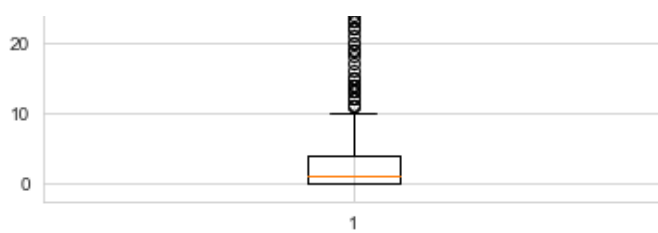
```
plt.boxplot("nodes",data=haberman)
#the 25% to 75% node lies between 0 to 6 and many outliers present
```

```
{'whiskers': [<matplotlib.lines.Line2D at 0x8ffaae9640>,
  <matplotlib.lines.Line2D at 0x8ffaae99a0>],
 'caps': [<matplotlib.lines.Line2D at 0x8ffaae9d00>,
  <matplotlib.lines.Line2D at 0x8ffaaf40a0>],
 'boxes': [<matplotlib.lines.Line2D at 0x8ffaae92e0>],
 'medians': [<matplotlib.lines.Line2D at 0x8ffaaf4400>],
 'fliers': [<matplotlib.lines.Line2D at 0x8ffaaf4760>],
 'means': []}
```
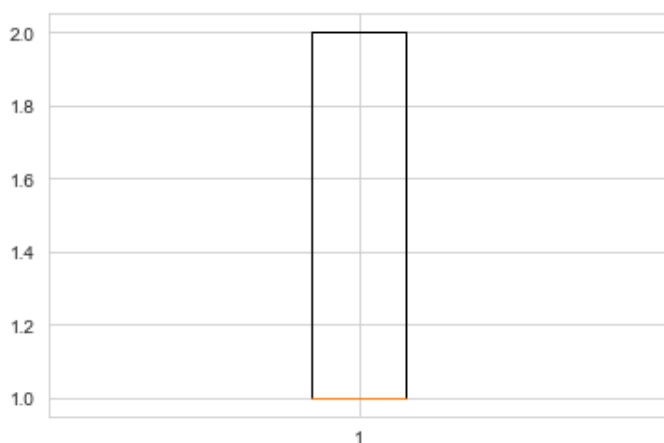
```
plt.boxplot("year",data=haberman)
#the 25% to 75% year lies between 60 to 66 and many outliers present
```

Out[57]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x8ffad49940>,
  <matplotlib.lines.Line2D at 0x8ffad49ca0>],
 'caps': [<matplotlib.lines.Line2D at 0x8ffad57040>,
  <matplotlib.lines.Line2D at 0x8ffad573a0>],
 'boxes': [<matplotlib.lines.Line2D at 0x8ffad495e0>],
 'medians': [<matplotlib.lines.Line2D at 0x8ffad57700>],
 'fliers': [<matplotlib.lines.Line2D at 0x8ffad57a60>],
 'means': []}
```


In [58]:

```
plt.boxplot("status",data=haberman)
```

Out[58]:

```
{'whiskers': [<matplotlib.lines.Line2D at 0x8ffad9dc10>,
  <matplotlib.lines.Line2D at 0x8ffad9df70>],
 'caps': [<matplotlib.lines.Line2D at 0x8ffadaa310>,
  <matplotlib.lines.Line2D at 0x8ffadaa670>],
 'boxes': [<matplotlib.lines.Line2D at 0x8ffad9d8b0>],
 'medians': [<matplotlib.lines.Line2D at 0x8ffadaa9d0>],
 'fliers': [<matplotlib.lines.Line2D at 0x8ffadaad30>],
 'means': []}
```


In [62]:

```
sns.FacetGrid(haberman, hue="status", size=5) \
    .map(sns.distplot, "age") \
    .add_legend();
plt.show();
```

In [64]:

```
sns.FacetGrid(haberman, hue="status", size=5) \
    .map(sns.distplot, "nodes") \
    .add_legend();
plt.show();
```

nodes

In [65]:

```python
sns.FacetGrid(haberman, hue="status", size=5) \
    .map(sns.distplot, "year") \
    .add_legend();
plt.show();
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\axisgrid.py:316: UserWarning: The `siz
e` parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2551: FutureWarning:
`distplot` is a deprecated function and will be removed in a future version. Please adapt
your code to use either `displot` (a figure-level function with similar flexibility) or `
histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2551: FutureWarning:
`distplot` is a deprecated function and will be removed in a future version. Please adapt
your code to use either `displot` (a figure-level function with similar flexibility) or `
histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```
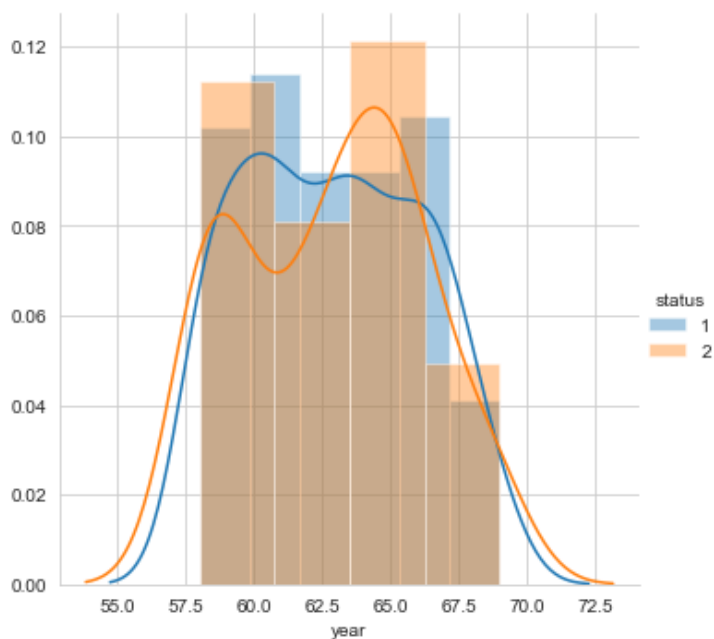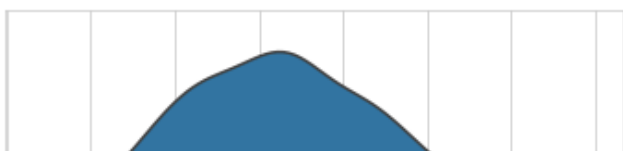

year

In [67]:
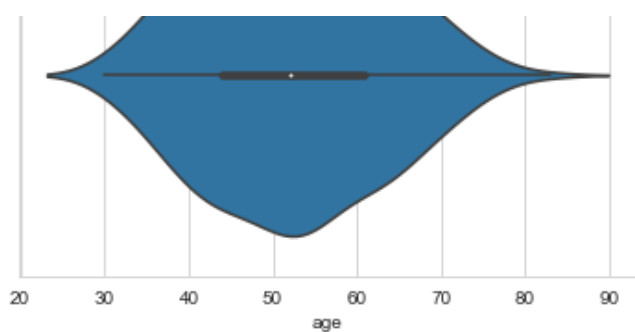
```python
sns.violinplot("age",data=haberman)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass
the following variable as a keyword arg: x. From version 0.12, the only valid positional
argument will be `data`, and passing other arguments without an explicit keyword will res
ult in an error or misinterpretation.
  warnings.warn(
```

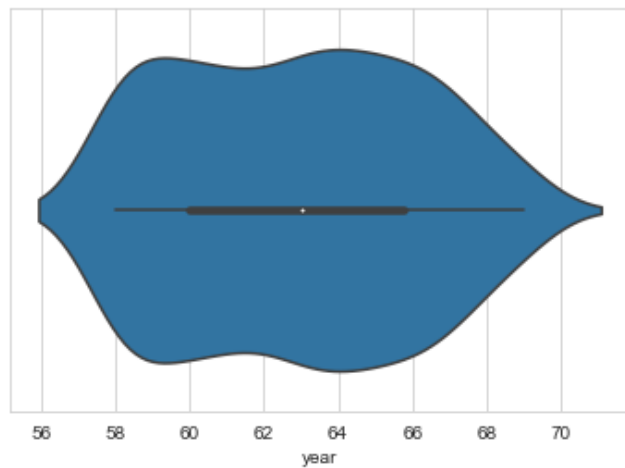Out[67]:

```
<AxesSubplot:xlabel='age'>
```

```
sns.violinplot("year",data=haberman)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass
the following variable as a keyword arg: x. From version 0.12, the only valid positional
argument will be `data`, and passing other arguments without an explicit keyword will res
ult in an error or misinterpretation.
  warnings.warn(
```
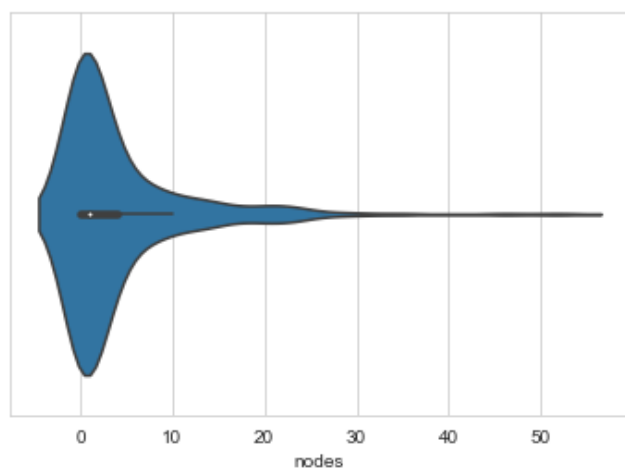
Out[68]:

```
<AxesSubplot:xlabel='year'>
```


In [69]:

```
sns.violinplot("nodes",data=haberman)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass
the following variable as a keyword arg: x. From version 0.12, the only valid positional
argument will be `data`, and passing other arguments without an explicit keyword will res
ult in an error or misinterpretation.
  warnings.warn(
```
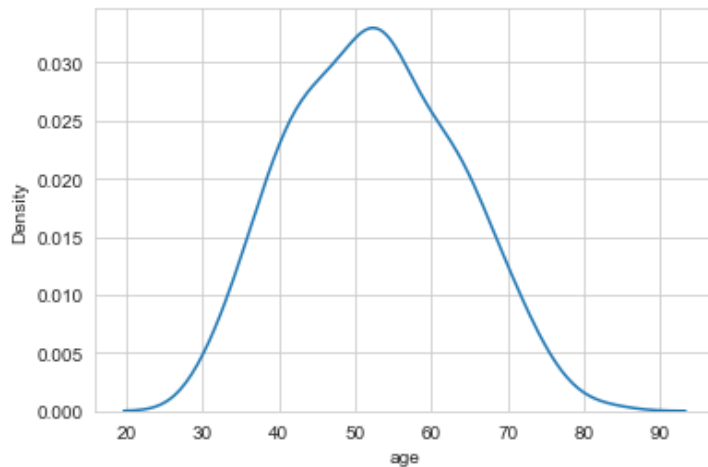
Out[69]:

```
<AxesSubplot:xlabel='nodes'>
```


In [86]:

```
sns.kdeplot("age",data=haberman)
```

Out[86]:

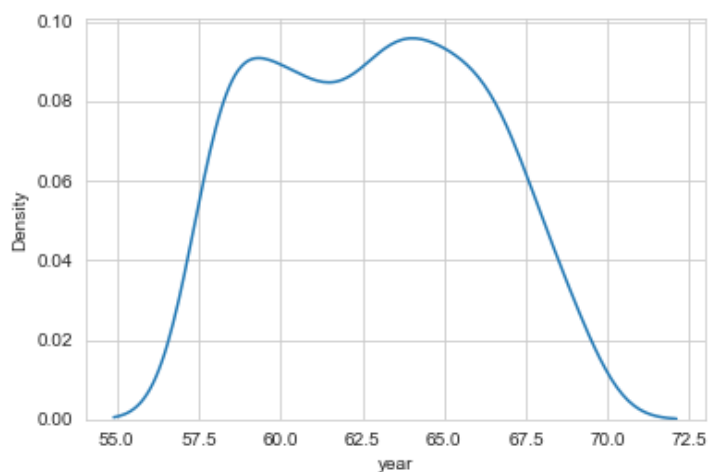<AxesSubplot:xlabel='age', ylabel='Density'>

In [87]:

```
sns.kdeplot("year",data=haberman)
```

Out[87]:

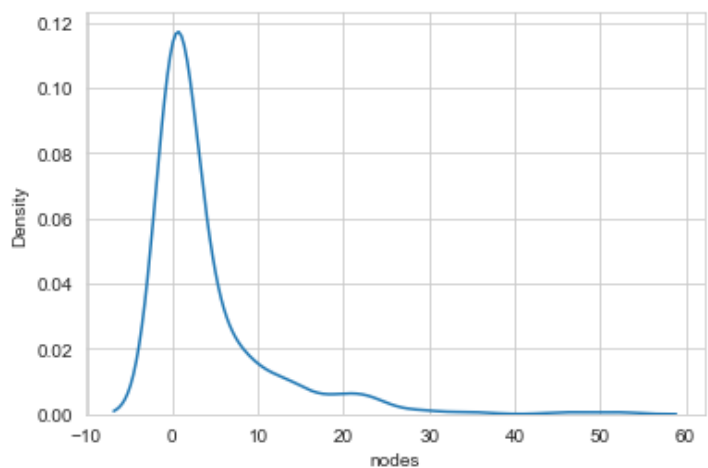<AxesSubplot:xlabel='year', ylabel='Density'>

In [88]:

```
sns.kdeplot("nodes",data=haberman)
#this distribution seems to be right squewed
```

Out[88]:

<AxesSubplot:xlabel='nodes', ylabel='Density'>
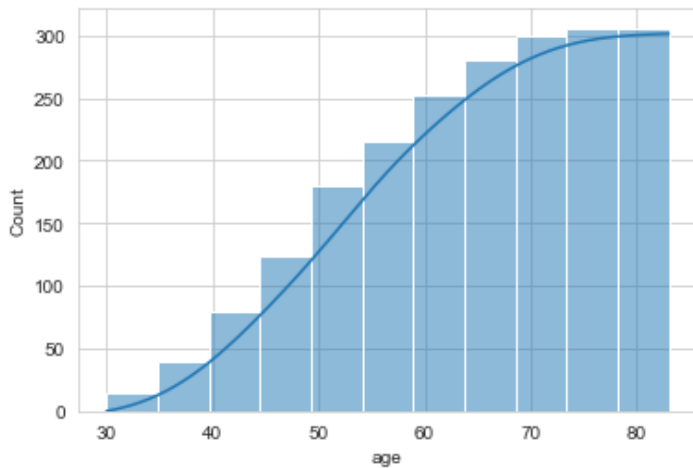
```
sns.histplot(data=haberman,x="age",cumulative=True,kde=True)
#cdf of age
```

Out[75]:

```
<AxesSubplot:xlabel='age', ylabel='Count'>
```
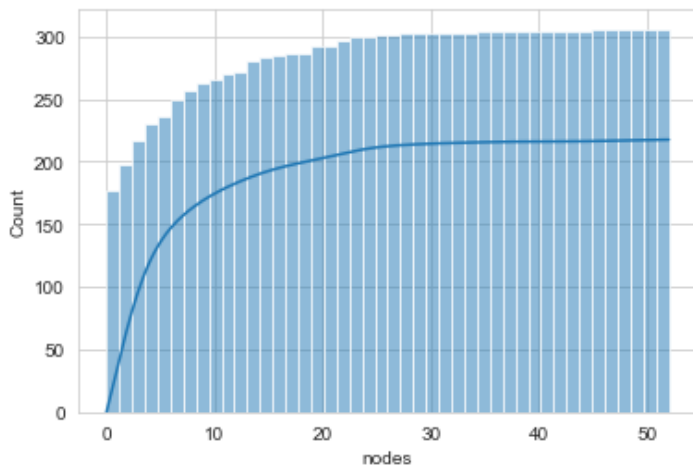

In [76]:

```
sns.histplot(data=haberman,x="nodes",cumulative=True,kde=True)
#cdf of nodes
```

Out[76]:

```
<AxesSubplot:xlabel='nodes', ylabel='Count'>
```


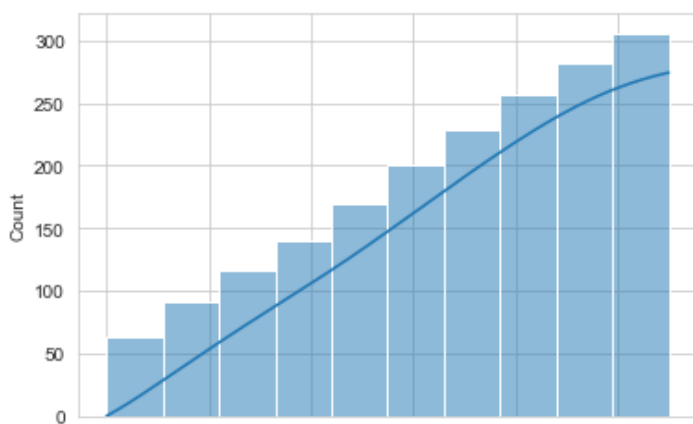In [79]:

```
sns.histplot(data=haberman,x="year",cumulative=True,kde=True)
#cdf of year
```

Out[79]:

```
<AxesSubplot:xlabel='year', ylabel='Count'>
```

58   60   62   64   66   68

year

In [84]:

```
null
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-84-d910b0287107> in <module>
----> 1 null

NameError: name 'null' is not defined
```