In [6]:

```
import os
import numpy as np
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt # Imported neccessary libraries
import scipy.stats as stats
```

In [7]:

```
data=pd.read_csv(r'C:\Users\anas.khanooni\Desktop\Assignment\insurance.csv')
```

In [11]:

```
data.head()
```

Out[11]:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

In [12]:

```
data.shape
```

Out[12]:

```
(1338, 7)
```

In [13]:

```
data.dtypes
```

Out[13]:

```
age           int64
sex          object
bmi         float64
children      int64
smoker       object
region       object
charges     float64
dtype: object
```

In [14]:

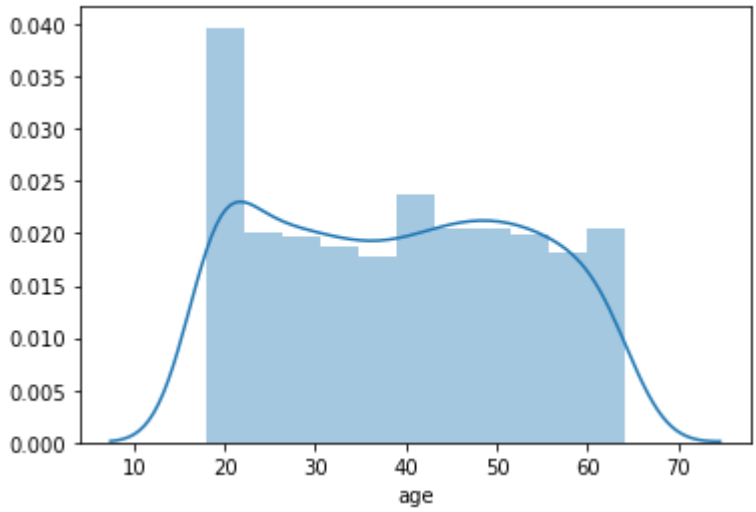
```
data.describe()
```

Out[14]:

	age	bmi	children	charges
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.663397	1.094918	13270.422265
std	14.049960	6.098187	1.205493	12110.011237
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000
75%	51.000000	34.693750	2.000000	16639.912515
max	64.000000	53.130000	5.000000	63770.428010

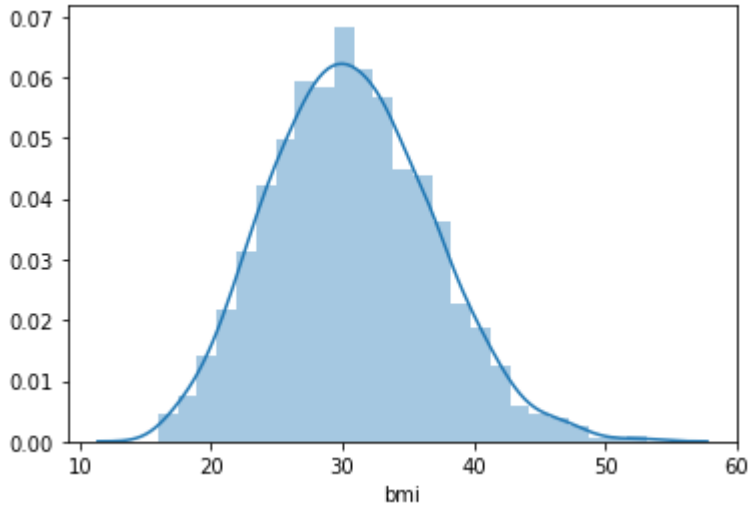
In [15]:

```
Histogram_Anas = sns.distplot(data['age'])
```



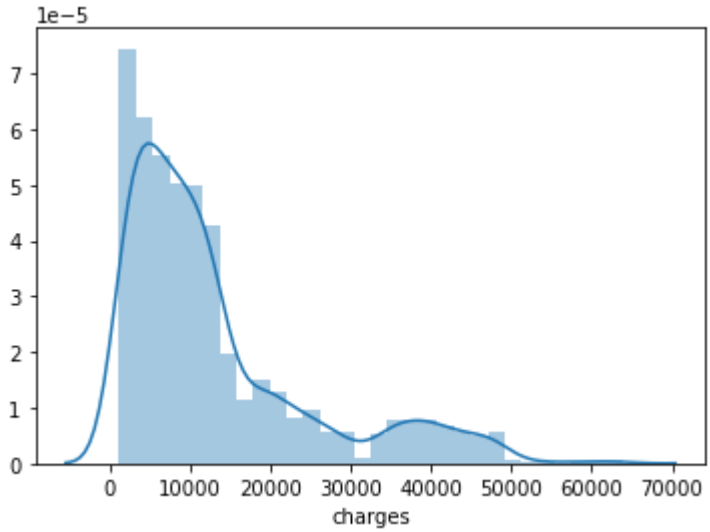
In [16]:

```
Histogram_Anas2 = sns.distplot(data['bmi'])
```



In [17]:

```
Histogram_Anas3 = sns.distplot(data['charges'])
```



In [18]:

```
data.skew(axis = 0, skipna = True)
```

Out[18]:

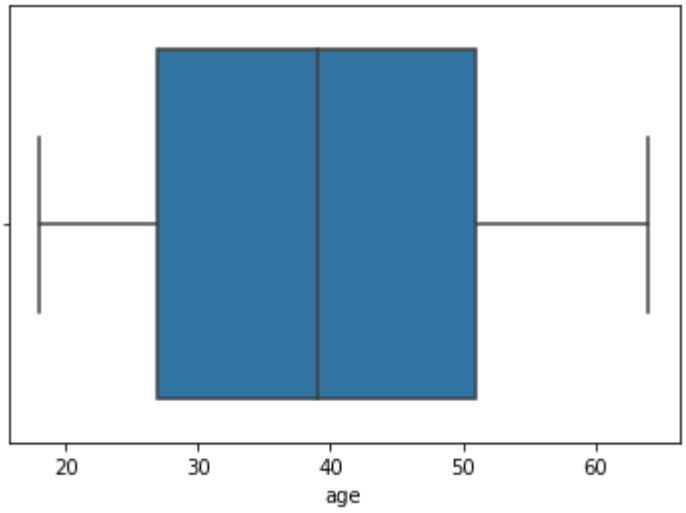
```
age         0.055673
bmi         0.284047
children    0.938380
charges     1.515880
dtype: float64
```

In [19]:

```
sns.boxplot(data['age'])
```

Out[19]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x19e29439b20>
```

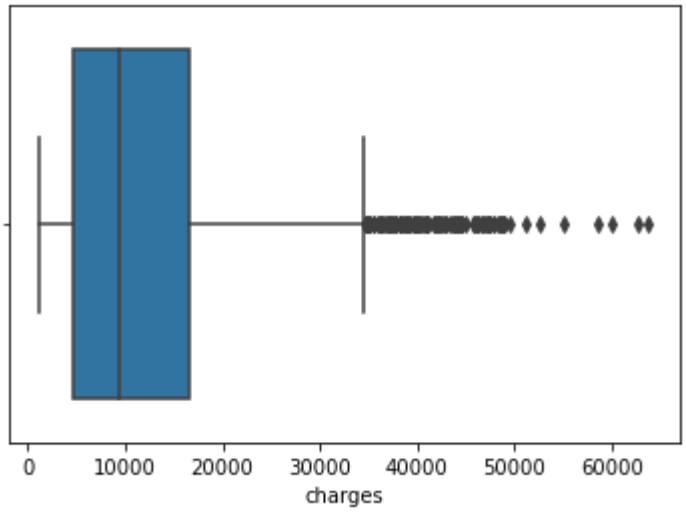


In [20]:

```
sns.boxplot(data['charges'])
```

Out[20]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x19e294950a0>
```

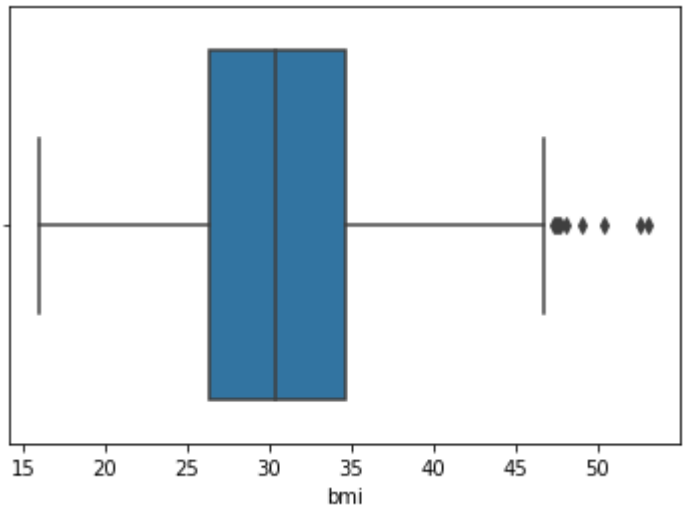


In [21]:

```
sns.boxplot(data['bmi'])
```

Out[21]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x19e294f58b0>
```

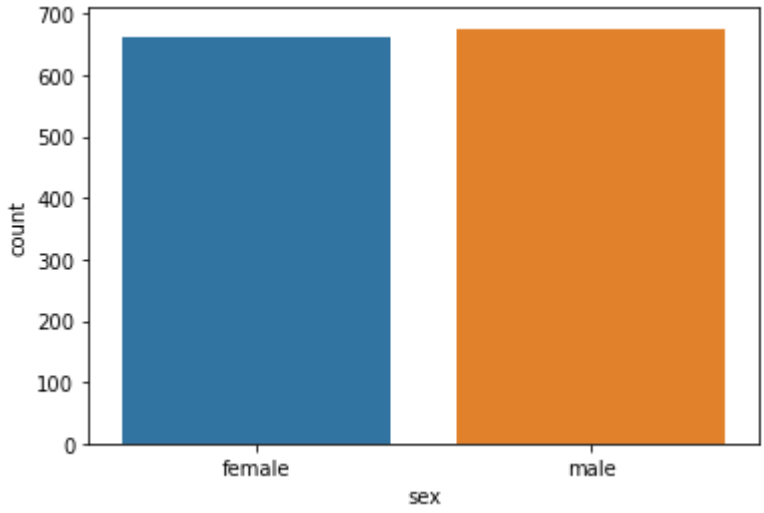


In [22]:

```
sns.countplot(data['sex'])
```

Out[22]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x19e29554e20>
```

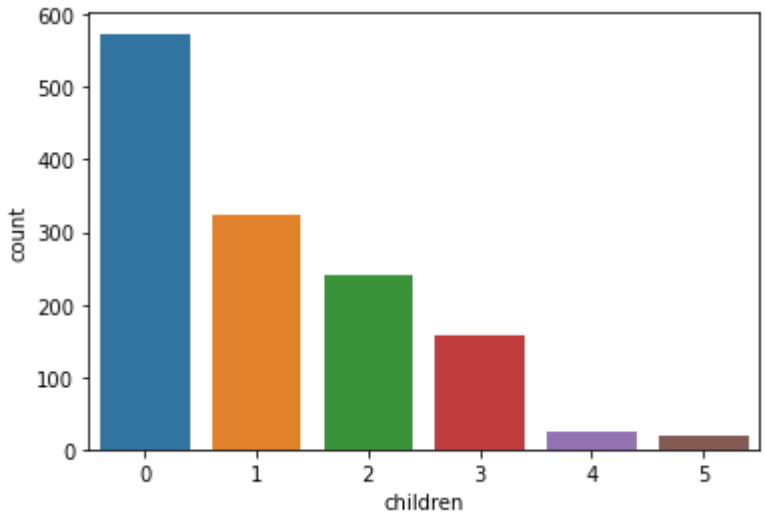


In [23]:

```
sns.countplot(data['children'])
```

Out[23]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x19e295b39a0>
```

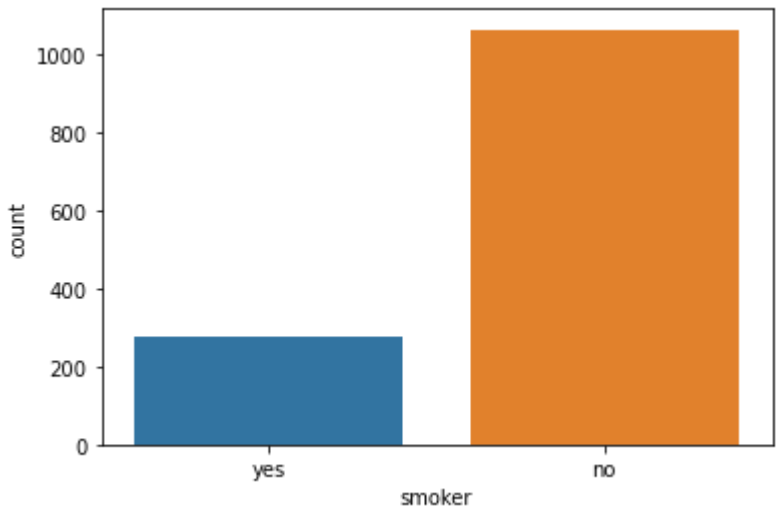


In [24]:

```
sns.countplot(data['smoker'])
```

Out[24]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x19e2960fbe0>
```

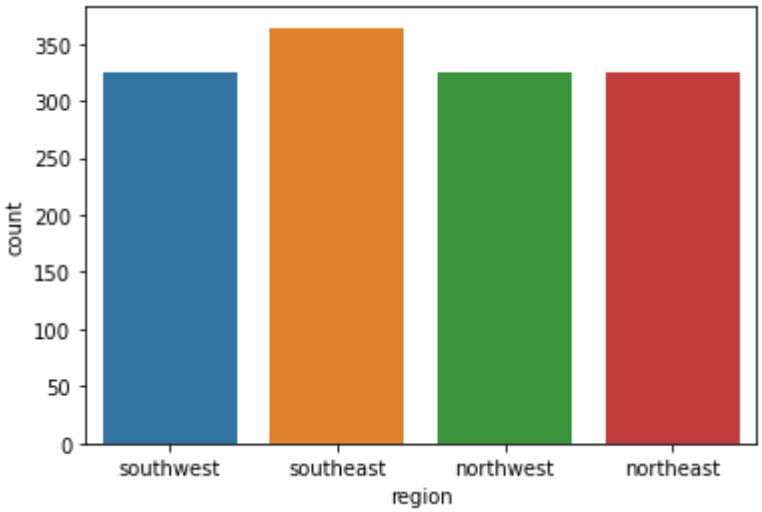


In [25]:

```
sns.countplot(data['region'])
```

Out[25]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x19e29646a00>
```

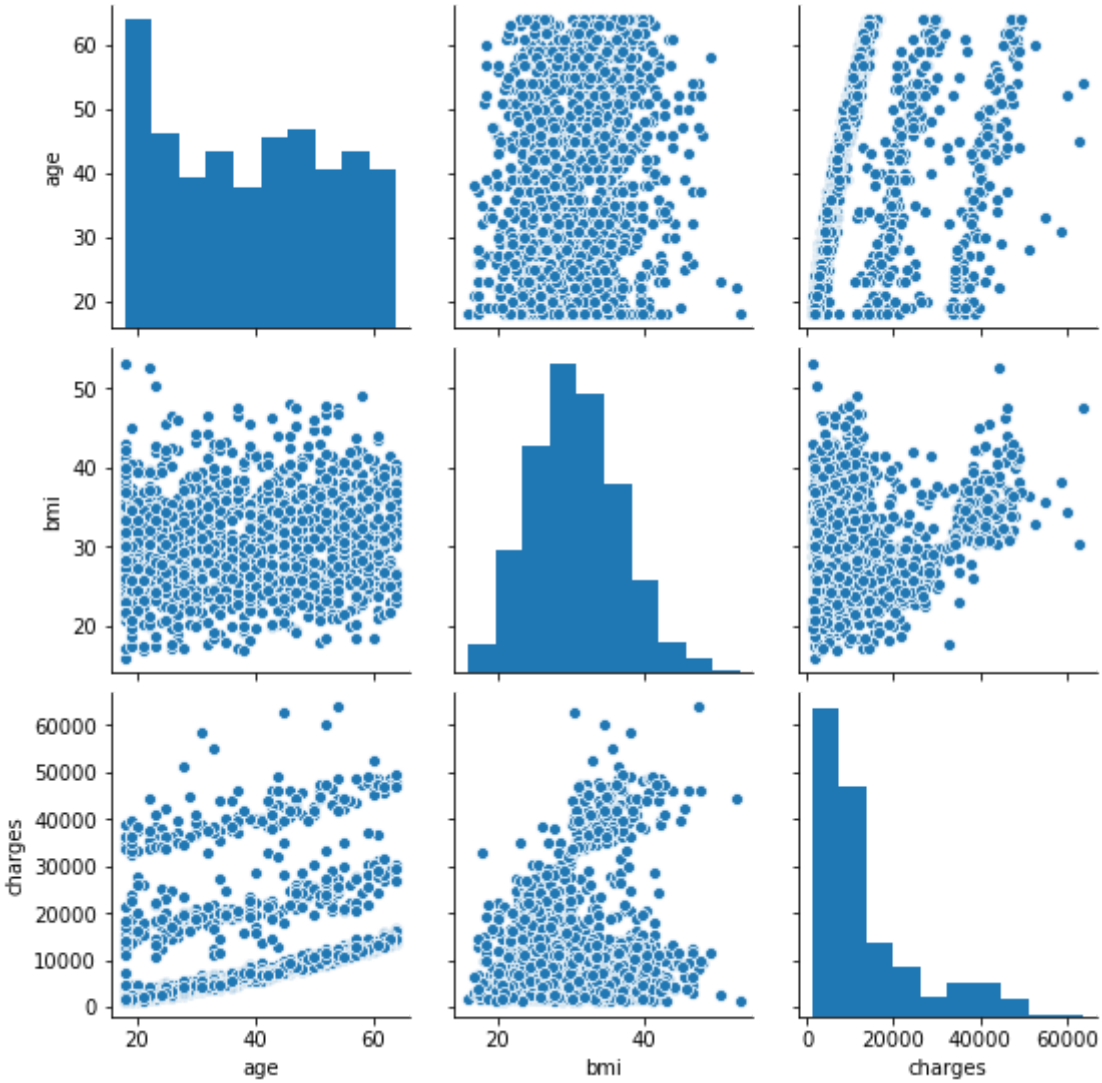


In [26]:

```
sns.pairplot(data[['age','bmi','charges']])
```

Out[26]:

```
<seaborn.axisgrid.PairGrid at 0x19e29694940>
```

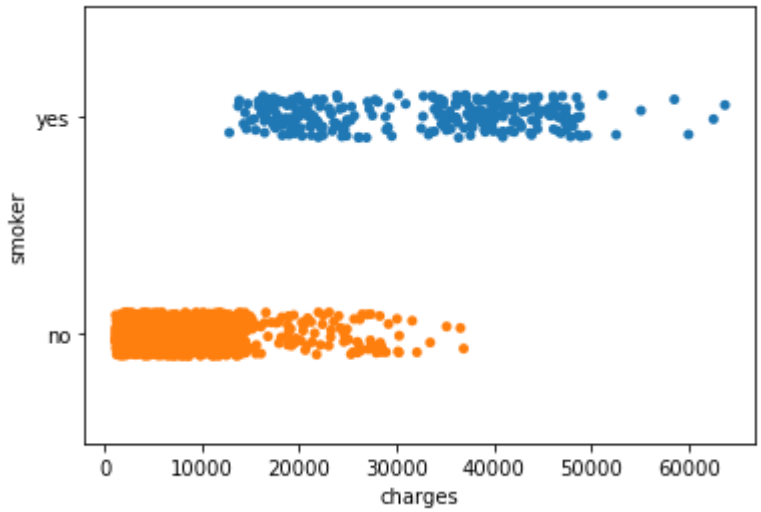

age
bmi
charges
20
30
40
50
60
0
10000
20000
30000
40000
50000
60000
20
40
60
0
20000
40000
60000

In [27]:

```
sns.stripplot(data['charges'], data['smoker']) #4a
```

Out[27]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x19e29b666d0>
```

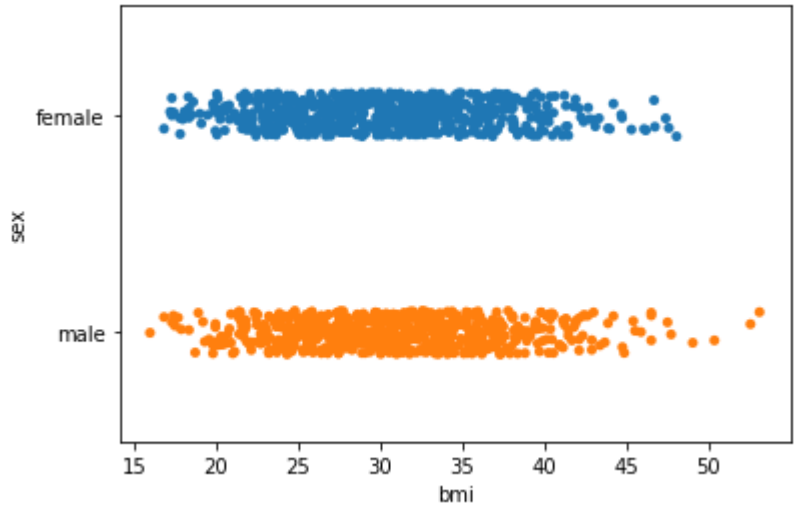


In [47]:

```
sns.stripplot(data['bmi'], data['sex']) #4b
```

Out[47]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x19e39c8e3a0>
```



In [48]:

```
print("Total count of smokers is ",data[data['smoker']=='yes'].shape[0]) #4c
print("Total count of male smokers is",data[data['smoker']=='yes'][data['sex']=='male']
.shape[0])
print("Total count of female smokers is",data[data['smoker']=='yes'][data['sex']=='fema
le'].shape[0])
print("Proportion of smokers who are male is",data[data['smoker']=='yes'][data['sex']==
'male'].shape[0]/data[data['smoker']=='yes'].shape[0])
print("Proportion of smokers who are female is", data[data[smokers]=='yes'][data['sex']
=='female'].shape[0]/data[data['smoker']=='yes'].shape[0])
```

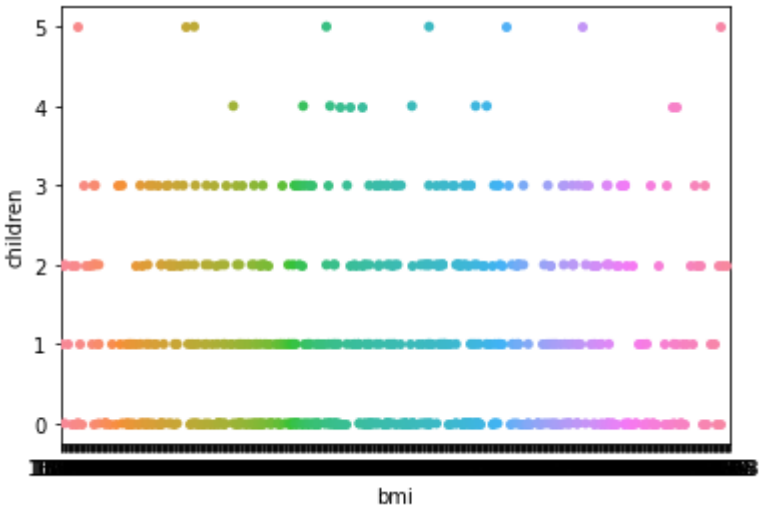
```
Total count of smokers is  274
Total count of male smokers is 159
Total count of female smokers is 115
Proportion of smokers who are male is 0.5802919708029197

<ipython-input-48-357e85068f67>:2: UserWarning: Boolean Series key will be
reindexed to match DataFrame index.
  print("Total count of male smokers is",data[data['smoker']=='yes'][data
['sex']=='male'].shape[0])
<ipython-input-48-357e85068f67>:3: UserWarning: Boolean Series key will be
reindexed to match DataFrame index.
  print("Total count of female smokers is",data[data['smoker']=='yes'][dat
a['sex']=='female'].shape[0])
<ipython-input-48-357e85068f67>:4: UserWarning: Boolean Series key will be
reindexed to match DataFrame index.
  print("Proportion of smokers who are male is",data[data['smoker']=='ye
s'][data['sex']=='male'].shape[0]/data[data['smoker']=='yes'].shape[0])

---------------------------------------------------------------------------
-
NameError                                 Traceback (most recent call las
t)
<ipython-input-48-357e85068f67> in <module>
      3 print("Total count of female smokers is",data[data['smoker']=='ye
s'][data['sex']=='female'].shape[0])
      4 print("Proportion of smokers who are male is",data[data['smoker']=
='yes'][data['sex']=='male'].shape[0]/data[data['smoker']=='yes'].shape[0]
)
----> 5 print("Proportion of smokers who are female is", data[data[smokers
]=='yes'][data['sex']=='female'].shape[0]/data[data['smoker']=='yes'].shap
e[0])

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

In [12]:

```
sns.stripplot(data['bmi'], data[data['sex']=='female']['children'])#4d
```

Out[12]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x193152b03a0>
```



In [ ]: