Karanjot Singh

In [82]:

```python
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv("haberman.csv")
df = pd.DataFrame(data)
print(df)
sns.pairplot(df,hue="Surv_status")
print(df.dtypes)
plt.show()
```

```
     Age  Op_Year  axil_nodes  Surv_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 x 4 columns]
Age            int64
Op_Year        int64
axil_nodes     int64
Surv_status    int64
dtype: object
```
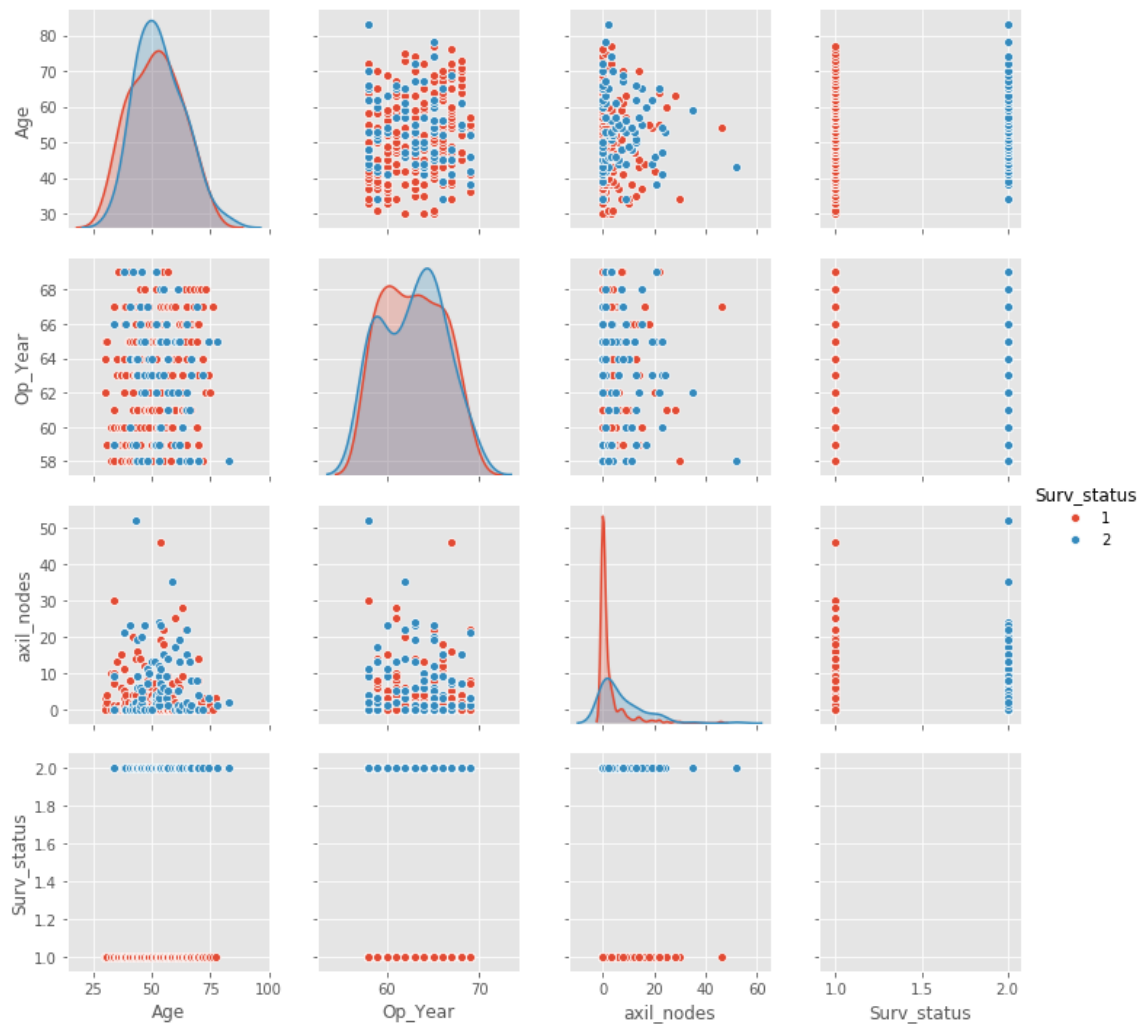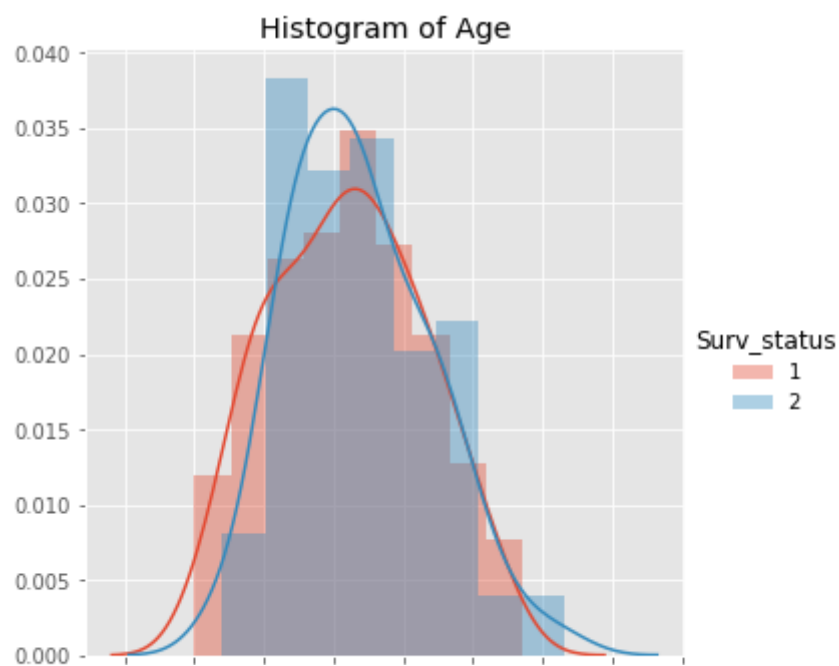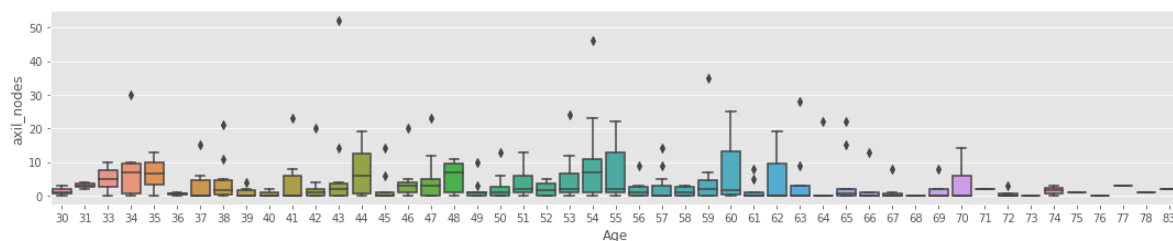
In [100]:

```
1  sns.FacetGrid(df,hue = 'Surv_status',size=5).map(sns.distplot,'Age').add_legend()
2  plt.title('Histogram of Age')
3  plt.show()
```

In [86]:

```
1  sns.FacetGrid(df,aspect=5)
2  sns.boxplot('Age','axil_nodes',data = df)
3  plt.show()
```
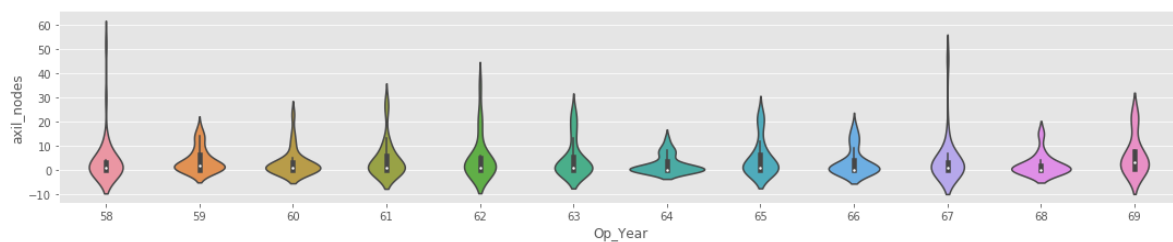

In [91]:

```
1  sns.FacetGrid(df,aspect=5)
2  sns.violinplot('Op_Year','axil_nodes',data = df)
3  #plt.show()
```

Out[91]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x233b58abd08>
```


In [101]:

```
1  df.describe()
```

Out[101]:

	Age	Op_Year	axil_nodes	Surv_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

In [102]:

```
1  df.median()
```

Out[102]:

```
Age            52.0
Op_Year        63.0
axil_nodes      1.0
Surv_status     1.0
dtype: float64
```

In [103]:

```
1  df.var()
```

Out[103]:

```
Age            116.714583
Op_Year         10.558631
axil_nodes      51.691118
Surv_status      0.195275
dtype: float64
```

In [107]:

```
1  print('Percentile')
2  df.quantile(0.1,axis = 0)
```

Percentile

Out[107]:

```
Age            38.0
Op_Year        58.0
axil_nodes      0.0
Surv_status     1.0
Name: 0.1, dtype: float64
```

In [105]:

```
1  df.quantile([.1,.25,.5,.75],axis=0)
```

Out[105]:

	Age	Op_Year	axil_nodes	Surv_status
0.10	38.00	58.00	0.0	1.0
0.25	44.00	60.00	0.0	1.0
0.50	52.00	63.00	1.0	1.0
0.75	60.75	65.75	4.0	2.0