

```
In [11]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

df=pd.read_csv("haberman.csv")
```

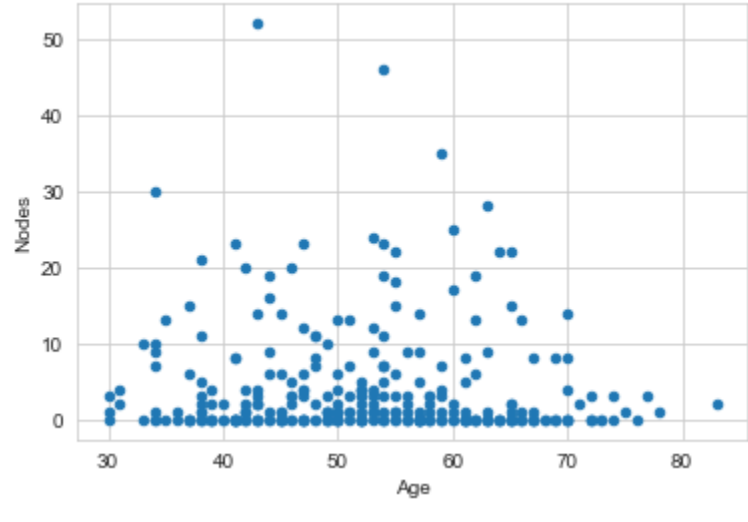
```
In [14]: print(df.shape)
print(df.columns)
df["Survival_Status"].value_counts()

(306, 4)
Index(['Age', 'Year', 'Nodes', 'Survival_Status'], dtype='object')
```

```
Out[14]: 1    225
2     81
Name: Survival_Status, dtype: int64
```

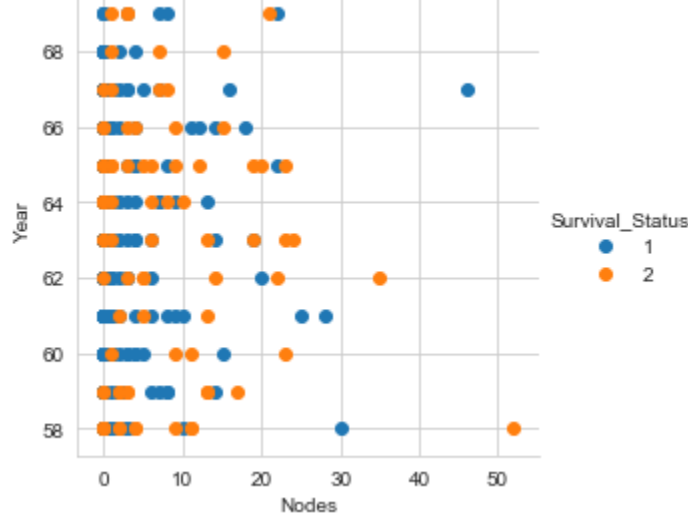
```
In [15]: df.plot(kind='scatter', x='Age', y='Nodes')

Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x1f6e6a00240>
```



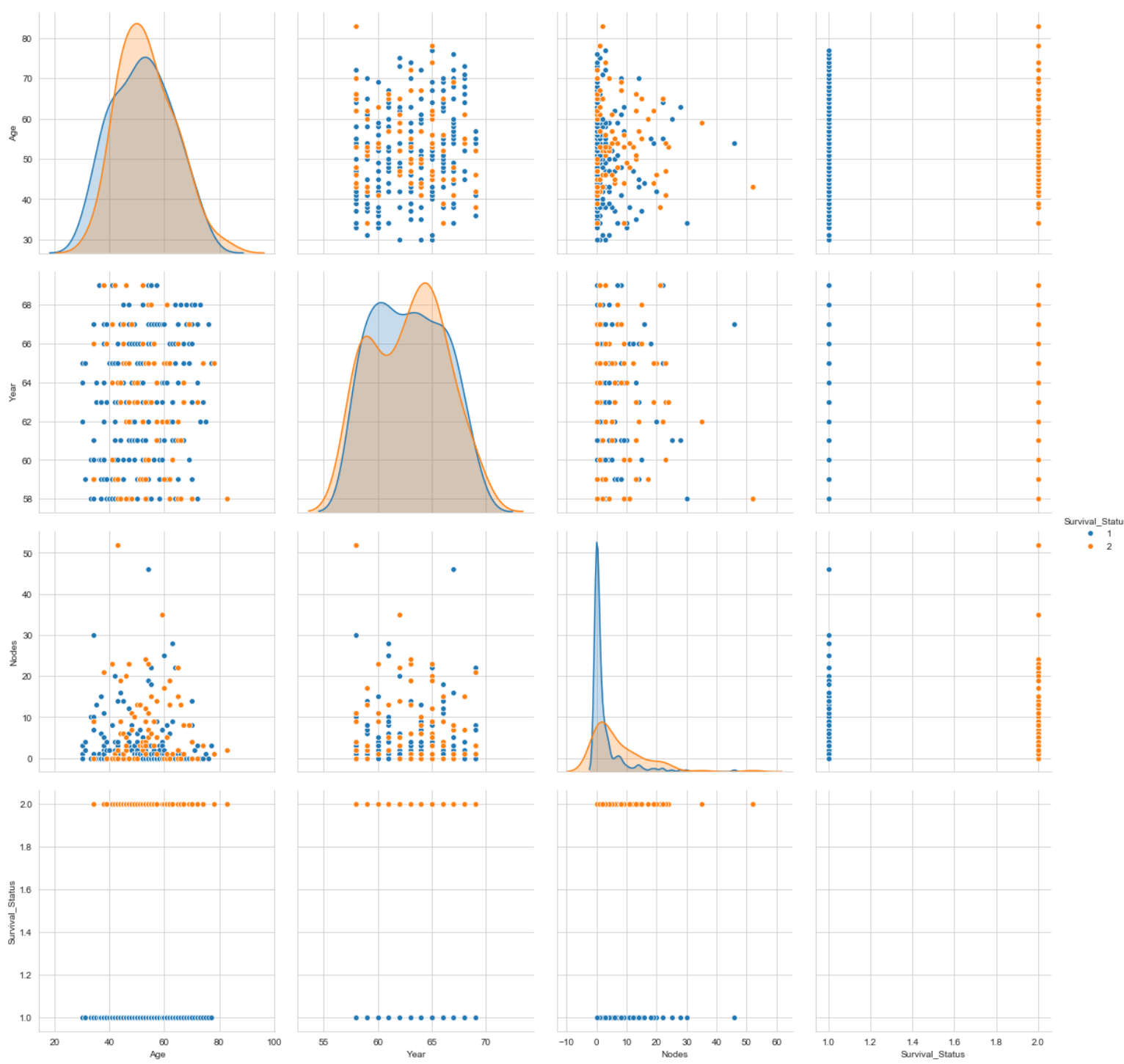
```
In [16]: sns.set_style("whitegrid");sns.FacetGrid(df, hue="Survival_Status", size=4) \
.map(plt.scatter, "Nodes","Year") \
.add_legend();
```

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\axisgrid.py:230: UserWarning: The 'size' parameter has been renamed to 'height'; please update your code.
warnings.warn(msg, UserWarning)



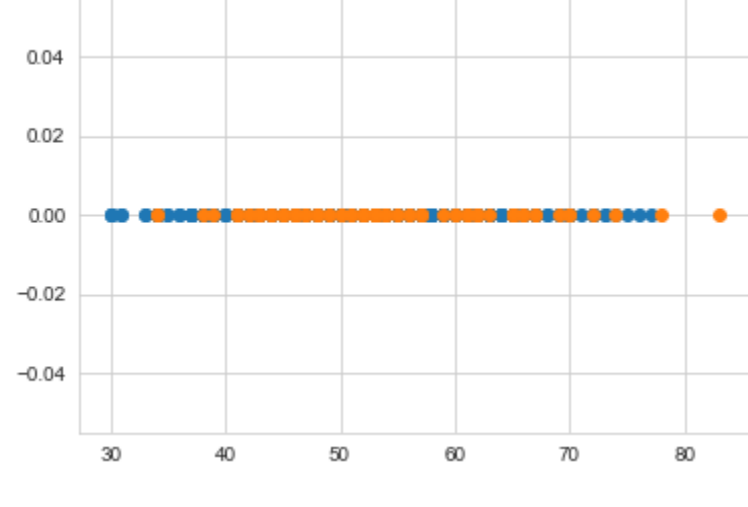
```
In [23]: plt.close();
sns.set_style("whitegrid");
sns.pairplot(df, hue="Survival_Status", size=4);
plt.show()
```

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\axisgrid.py:2065: UserWarning: The 'size' parameter has been renamed to 'height'; please update your code.
warnings.warn(msg, UserWarning)



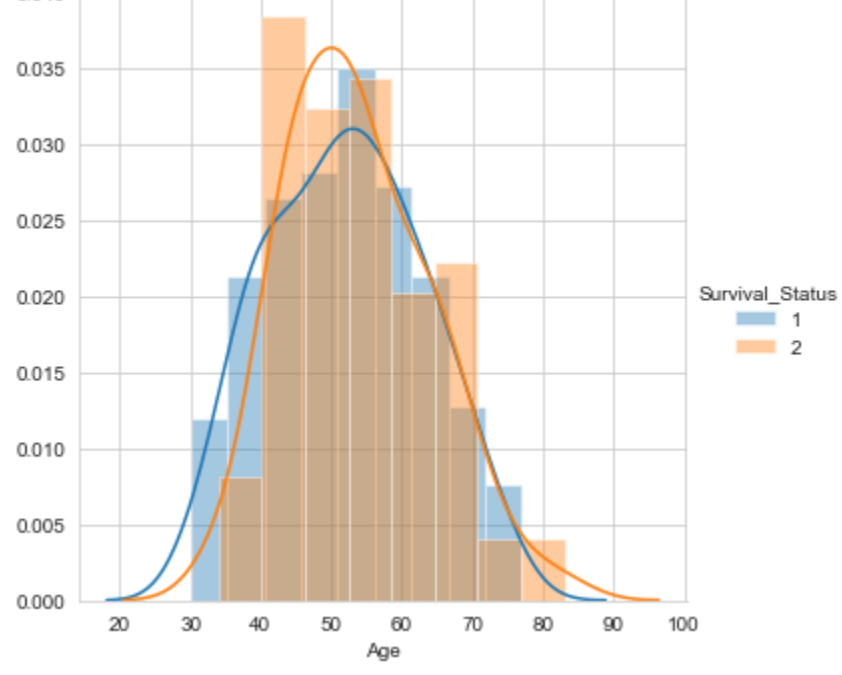
```
In [22]: import numpy as np
df_one = df.loc[df["Survival_Status"] == 1];
df_two = df.loc[df["Survival_Status"] == 2];
# iris_virginica = iris.loc[iris["species"] == "virginica"];
# iris_versicolor = iris.loc[iris["species"] == "versicolor"];
#print(iris_setosa["petal_length"])
plt.plot(df_one["Age"], np.zeros_like(df_one['Age']), 'o')
plt.plot(df_two["Age"], np.zeros_like(df_two['Age']), 'o')
# plt.plot(iris_versicolor["petal_length"], np.zeros_like(iris_versicolor['petal_length']), 'o')
# plt.plot(iris_virginica["petal_length"], np.zeros_like(iris_virginica['petal_length']), 'o')

plt.show()
```



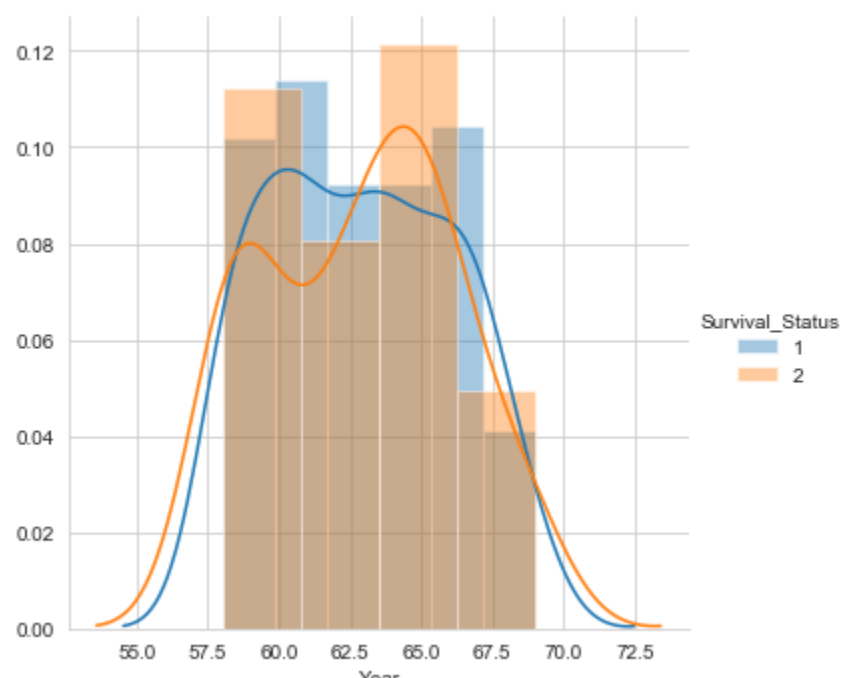
```
In [24]: sns.FacetGrid(df, hue="Survival_Status", size=5) \
.map(sns.distplot, "Age") \
.add_legend();
plt.show();
```

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\axisgrid.py:230: UserWarning: The 'size' parameter has been renamed to 'height'; please update your code.
warnings.warn(msg, UserWarning)



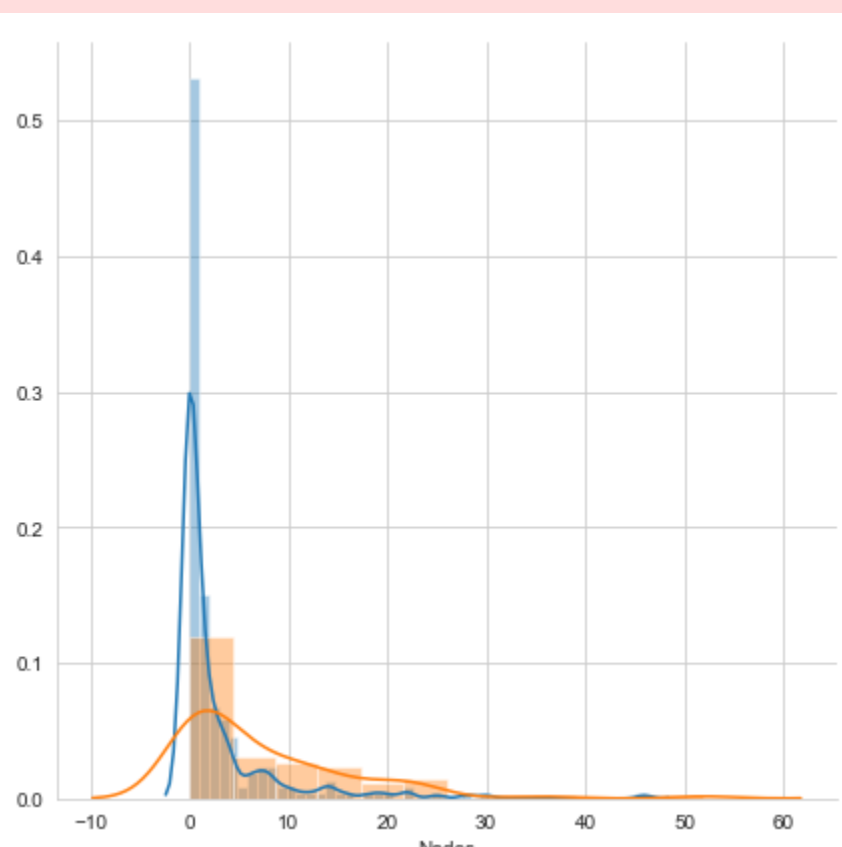
```
In [25]: sns.FacetGrid(df, hue="Survival_Status", size=5) \
.map(sns.distplot, "Year") \
.add_legend();
plt.show();
```

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\axisgrid.py:230: UserWarning: The 'size' parameter has been renamed to 'height'; please update your code.
warnings.warn(msg, UserWarning)



```
In [28]: Sns.FacetGrid(df, hue="Survival_Status", size=6) \
.map(sns.distplot, "Nodes") \
.add_legend();
plt.show();
```

C:\ProgramData\Anaconda3\lib\site-packages\seaborn\axisgrid.py:230: UserWarning: The 'size' parameter has been renamed to 'height'; please update your code.
warnings.warn(msg, UserWarning)



```
In [31]: counts, bin_edges = np.histogram(df_one['Age'], bins=10,
density = True)

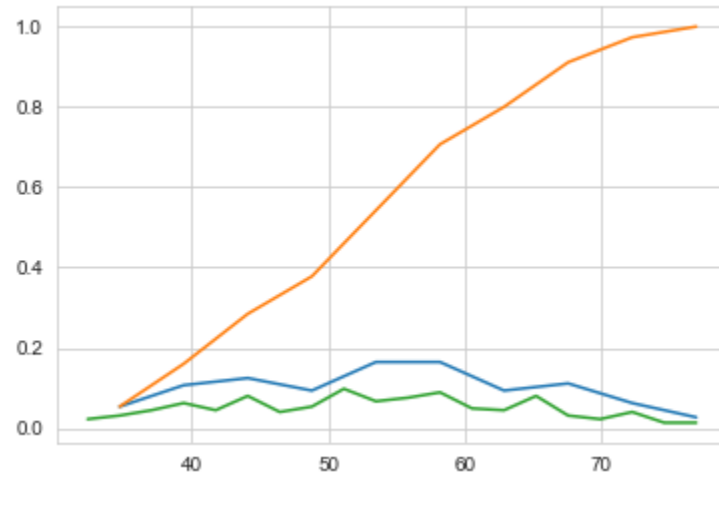
pdf = counts/(sum(counts))
print(pdf);
print(bin_edges);
cdf = np.cumsum(pdf);
plt.plot(bin_edges[1:],pdf);
plt.plot(bin_edges[1:], cdf)

counts, bin_edges = np.histogram(df_one['Age'], bins=20,
density = True)

pdf = counts/(sum(counts))
plt.plot(bin_edges[1:],pdf);

plt.show();

[0.05333333 0.10666667 0.12444444 0.09333333 0.16444444 0.16444444
0.09333333 0.11111111 0.06222222 0.02666667]
[30.  34.7 39.4 44.1 48.8 53.5 58.2 62.9 67.6 72.3 77. ]
```



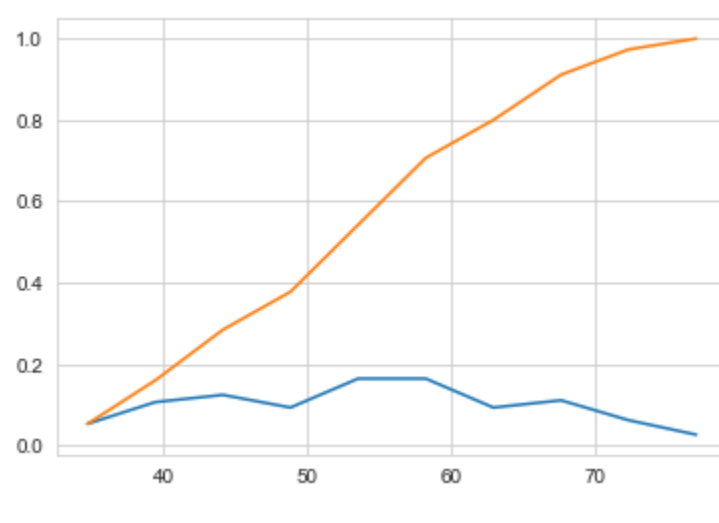
```
In [32]: counts, bin_edges = np.histogram(df_one['Age'], bins=10,
density = True)

pdf = counts/(sum(counts))
print(pdf);
print(bin_edges)

#compute CDF
cdf = np.cumsum(pdf)
plt.plot(bin_edges[1:],pdf)
plt.plot(bin_edges[1:], cdf)

plt.show();

[0.05333333 0.10666667 0.12444444 0.09333333 0.16444444 0.16444444
0.09333333 0.11111111 0.06222222 0.02666667]
[30.  34.7 39.4 44.1 48.8 53.5 58.2 62.9 67.6 72.3 77. ]
```



```
In [33]: counts, bin_edges = np.histogram(df_one['Age'], bins=10,
density = True)

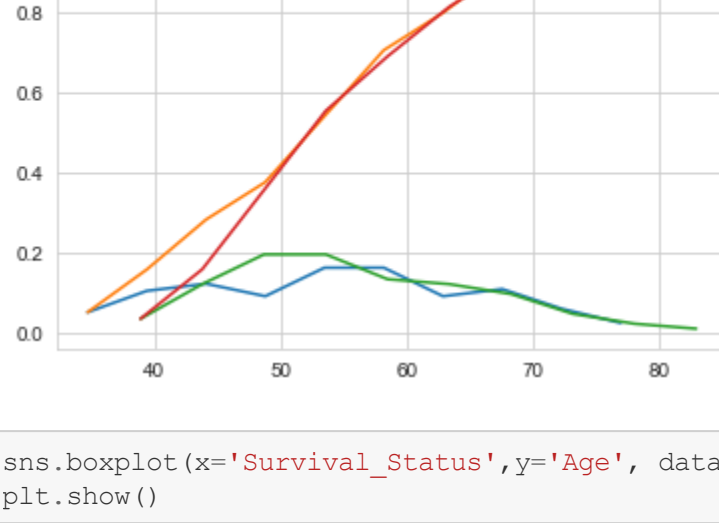
pdf = counts/(sum(counts))
print(pdf);
print(bin_edges)
cdf = np.cumsum(pdf)
plt.plot(bin_edges[1:],pdf)
plt.plot(bin_edges[1:], cdf)

# two
counts, bin_edges = np.histogram(df_two['Age'], bins=10,
density = True)

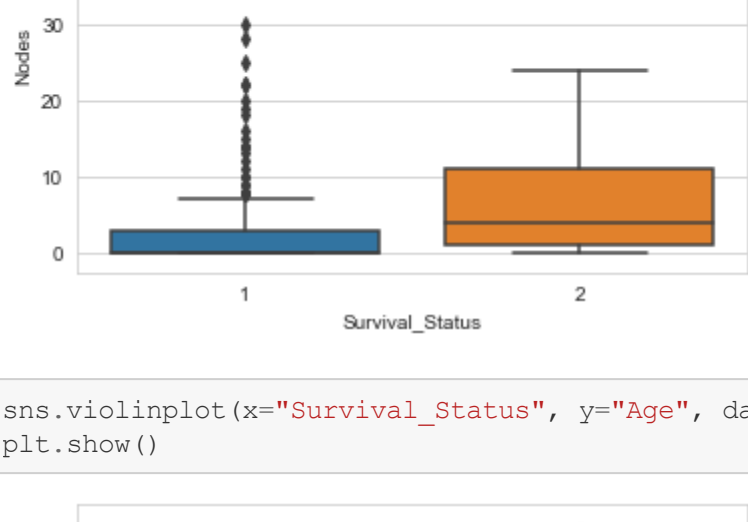
pdf = counts/(sum(counts))
print(pdf);
print(bin_edges)
cdf = np.cumsum(pdf)
plt.plot(bin_edges[1:],pdf)
plt.plot(bin_edges[1:], cdf)

[0.05333333 0.10666667 0.12444444 0.09333333 0.16444444 0.16444444
0.09333333 0.11111111 0.06222222 0.02666667]
[30.  34.7 39.4 44.1 48.8 53.5 58.2 62.9 67.6 72.3 77. ]
[0.03703704 0.12345679 0.19753086 0.19753086 0.13580247 0.12345679
0.09876543 0.04938272 0.02469136 0.01234568]
[34.  38.9 43.8 48.7 53.6 58.5 63.4 68.3 73.2 78.1 83. ]
```

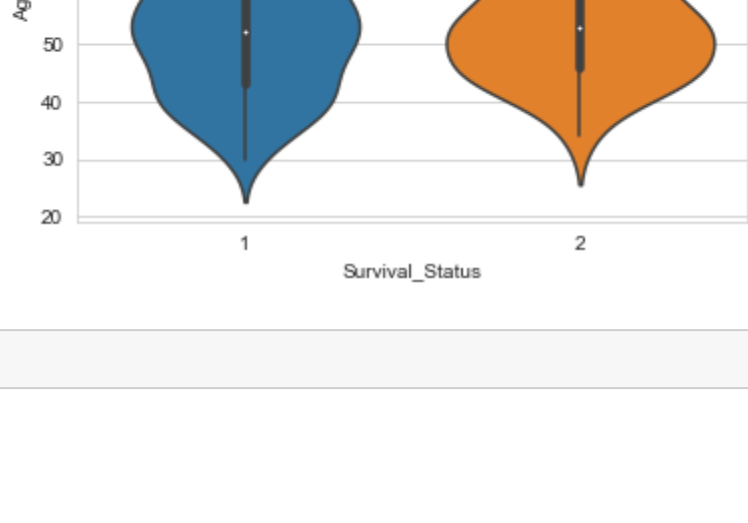
```
Out[33]: <matplotlib.lines.Line2D at 0x1f6e6263908>
```



```
In [39]: sns.boxplot(x='Survival_Status',y='Age', data=df)
plt.show()
```



```
In [37]: sns.violinplot(x="Survival_Status", y="Age", data=df, size=5)
plt.show()
```



```
In [ ]:
```