Exercise 08

This Jupyter Notebook is my solution to Exercise 08 It is a written analysis of the given dataset of titanic passengers.

In [1]:
```
# Prepare an analysis of the titanic dataset (data/titanic.csv).
# You are interested in determining factors that lead to death or survival of passengers.
# However, before actually training an ML model to predict death or survival, you should make
# yourself familiar with the data.
#
# Look into which features are available, what types of features these are, whether
# values are missing and how to deal with them, etc.
# You can then continue to create some visualizations of things you think might be interesting or relevant
# in predicting survival, or just generally useful to know about the dataset.
# The output should not just be prints of numbers or figures, but also some explanatory text of
# what you have been analysing, and which conclusions you can draw from the different steps of your analysis.
#
# You can do all of this in a Jupyter Notebook, embedding analysis, code and explanations in a single document.
# Alternatively you can create a separate script for the analysis, and then list results, figures and explanations
# either in the PR directly, or in some form of pdf/word/markdown/... document.
#
# For this exercise I very much encourage you to work together to come up with ideas of which
# calculations/plots/... might be interesting to analyse. You will find very many such analyses online
# for the titanic data, but I would recommend trying to think of your own ideas first, and only later
# look for more inspiration online.
```

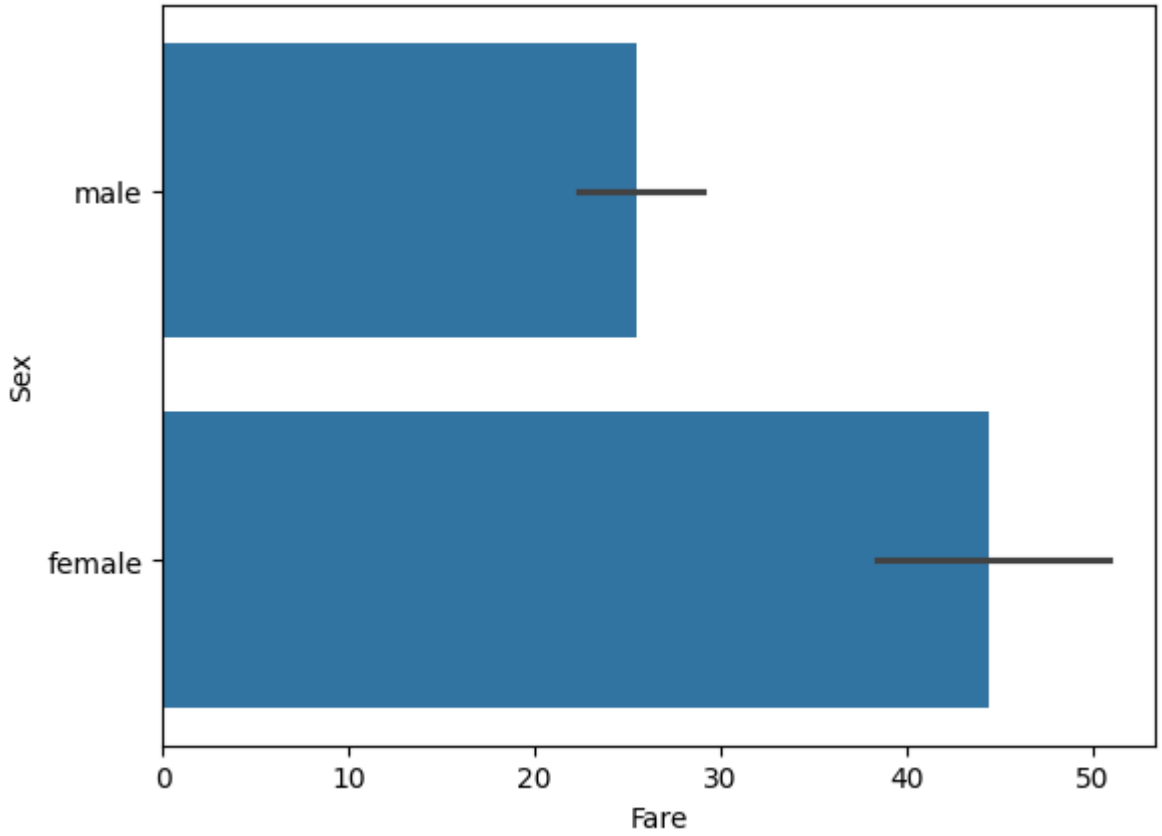
In [2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
```

In [3]:
```
df = pd.read_csv('Data/titanic.csv')
#df
```

In [11]:
```
sns.barplot(
    data=df,
    x = "Fare",
```

```
    y = "Sex",
)
```

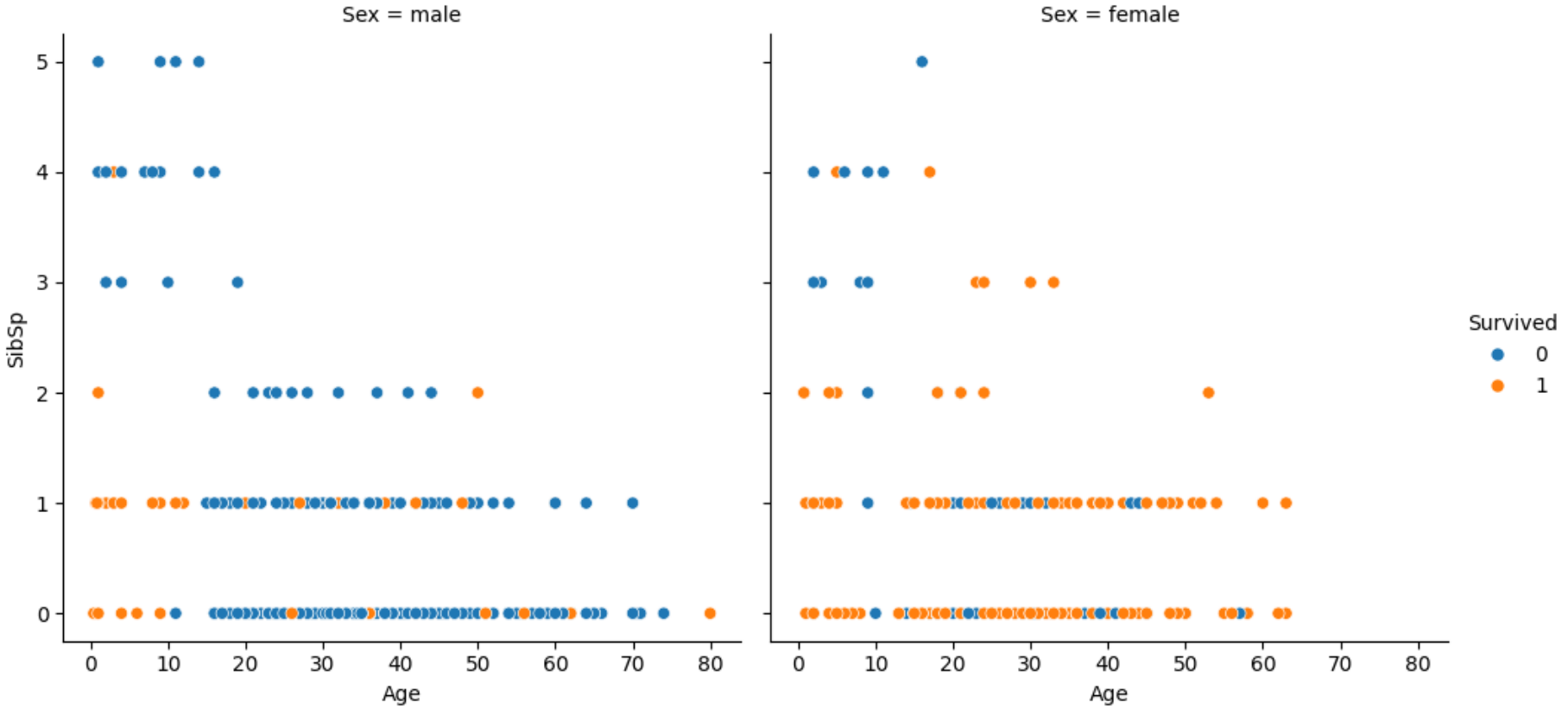
Out[11]: <Axes: xlabel='Fare', ylabel='Sex'>



In this graph we can see that women paid a higher Fare than men. This is caused by more men being in the third class and therefore paying the least fare

In [10]:
```
sns.relplot(
    data=df,
    x = "Age",
    y = "SibSp",
    hue = "Survived",
    col = "Sex")
```

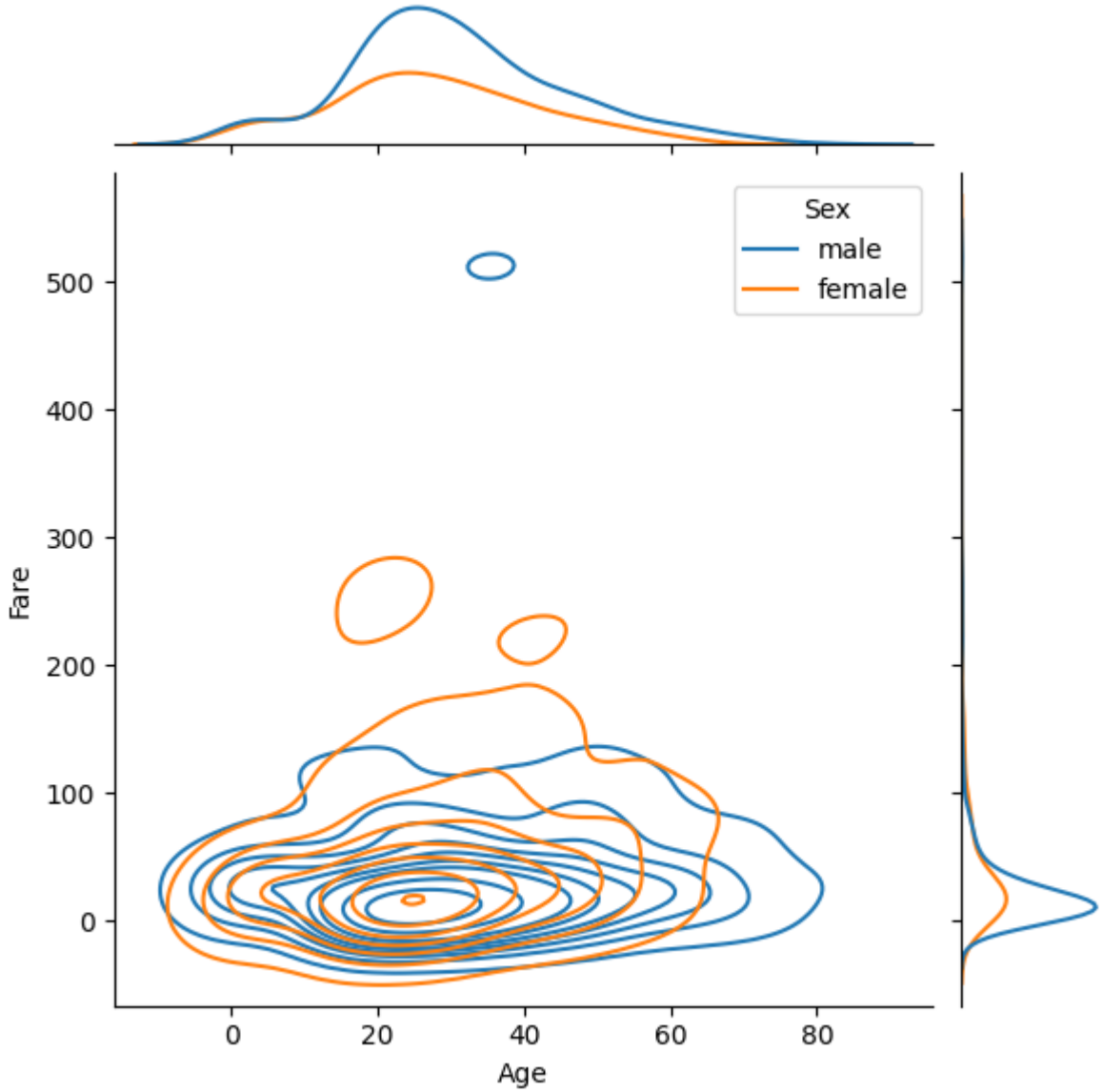
Out[10]: <seaborn.axisgrid.FacetGrid at 0x7f7ec5378fe0>



This graph shows us that there were more boys/men with a higher amount of siblings onboard which did not survive. Women/girls with low amount of siblings had the highest amount of survivors

In [6]:
```
sns.jointplot(data=df,
                x="Age",
                y="Fare",
                hue="Sex",
                #col="Survived",
                kind="kde"
              )
```

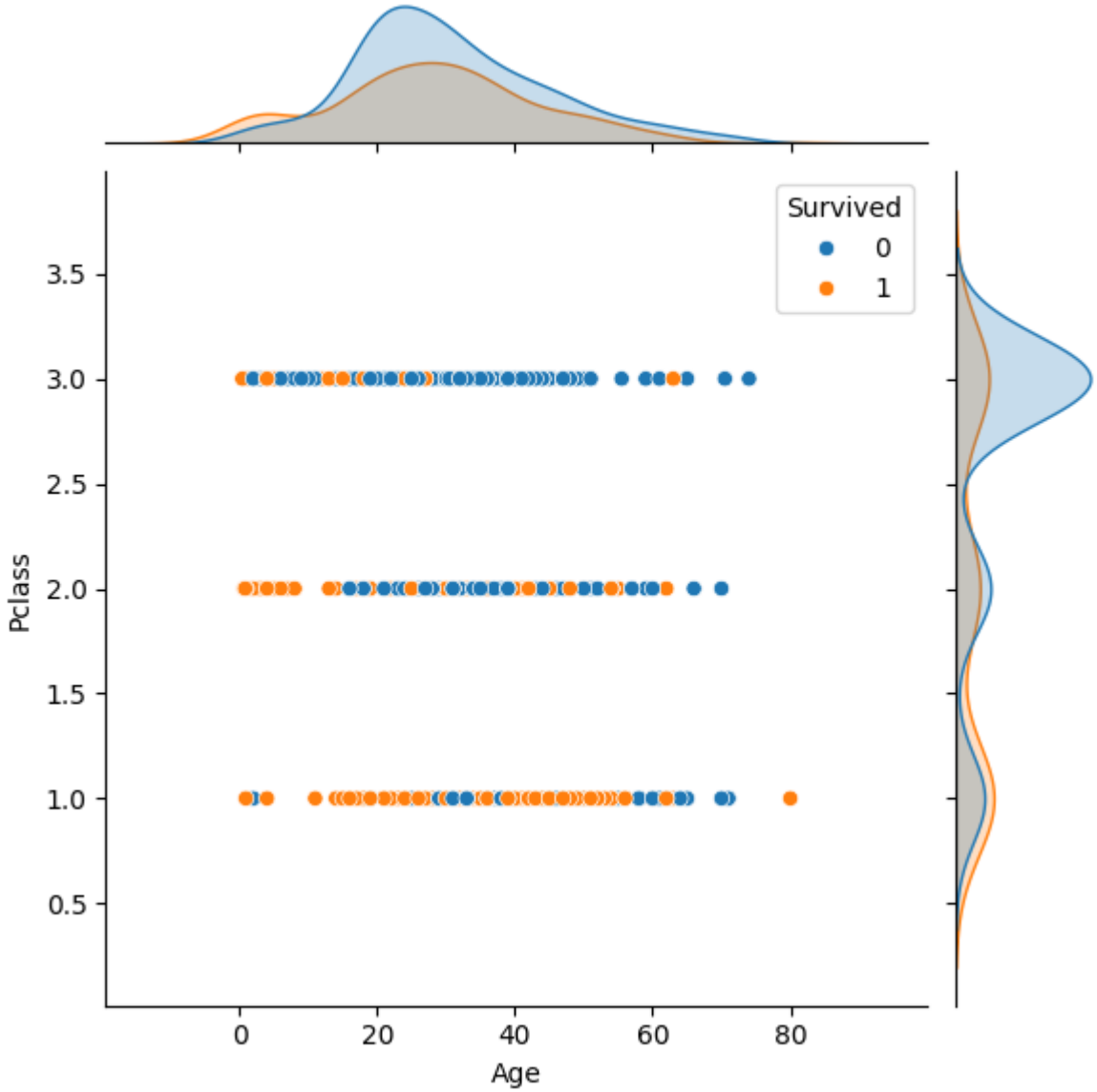
Out[6]: <seaborn.axisgrid.JointGrid at 0x7f7ec53794c0>



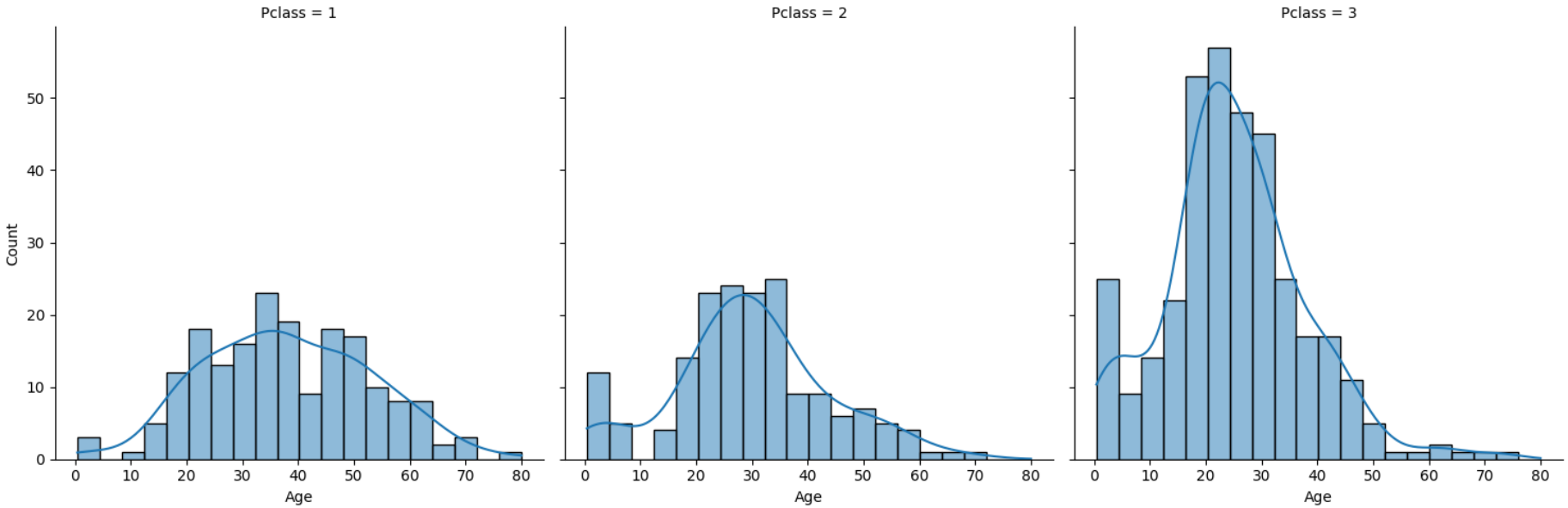
In [7]:
```
sns.jointplot(data=df,
              x="Age",
              y="Pclass",
              hue="Survived",
              #col="Survived",
```

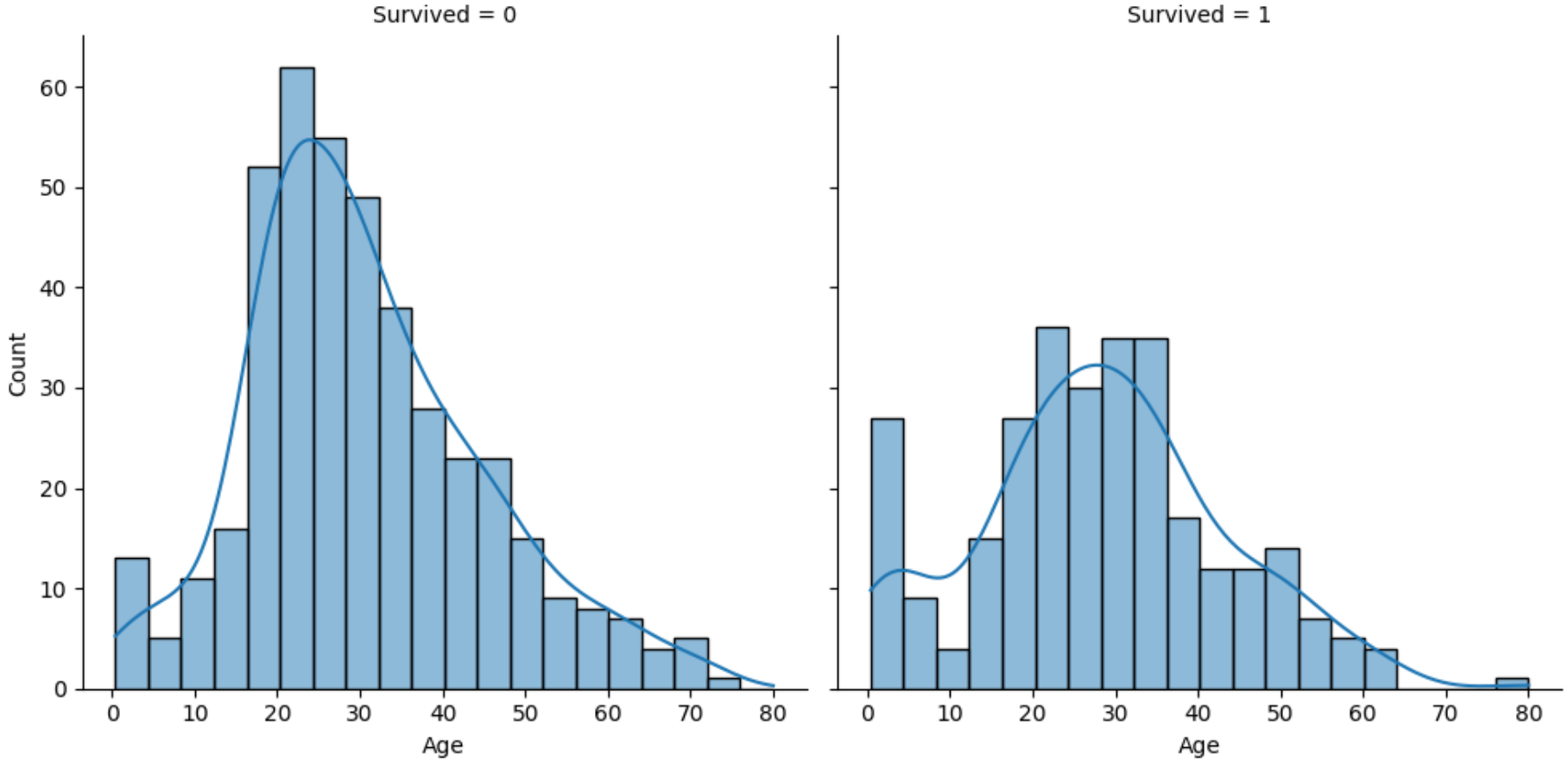
```
                kind="scatter"
)
```

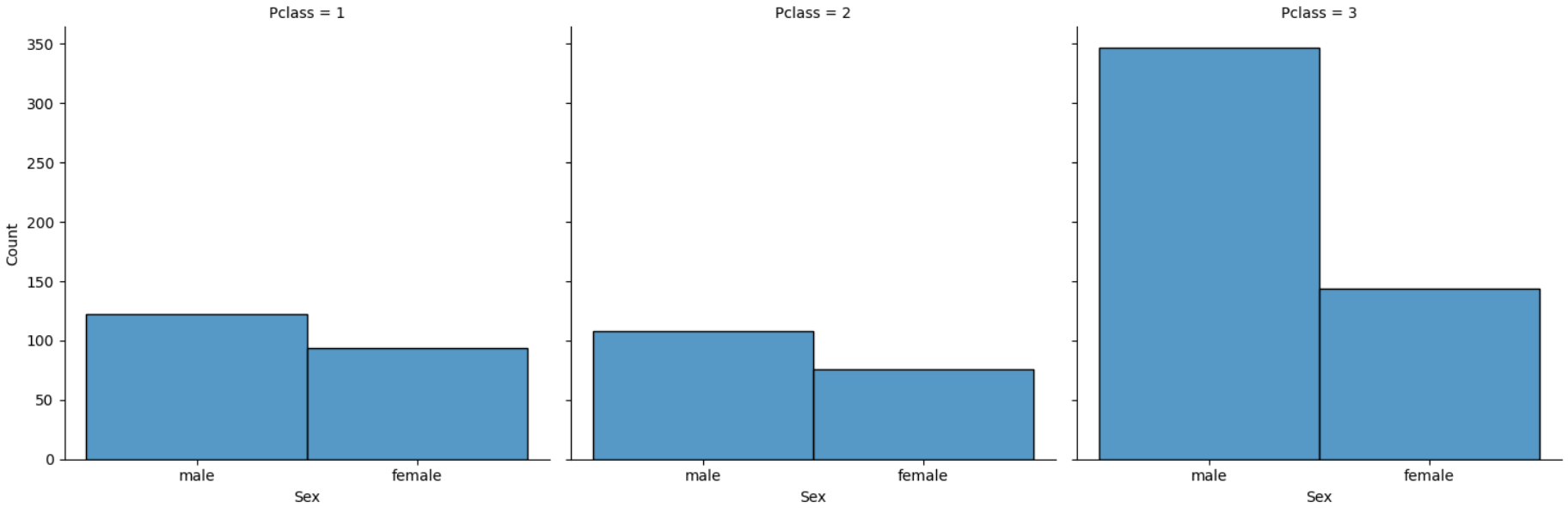
Out[7]: <seaborn.axisgrid.JointGrid at 0x7f7ec53c2270>



In [12]:
```
sns.displot(data=df, x="Age", col="Pclass", kde=True);
sns.displot(data=df, x="Age", col="Survived", kde=True);
sns.displot(data=df, x="Sex", col="Pclass", kde=False);
```



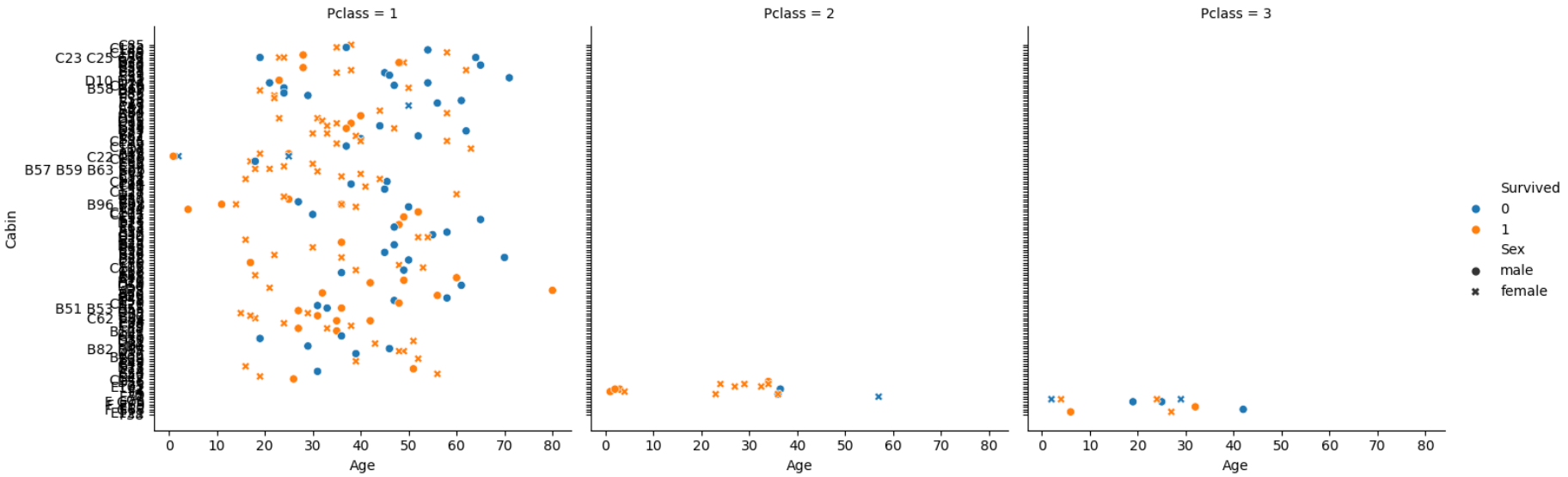

Survived = 0
Survived = 1
Count
Age
Age
0
10
20
30
40
50
60
70
80



Men dominated the third class while the first class and second class was not dominated by each of both genders. Most men had an age of about 25years in third class

In [9]:
```
sns.relplot(
    data=df,
    x = "Age",
    y = "Cabin",
    hue = "Survived",
    col="Pclass", style = "Sex")
```

Out[9]: <seaborn.axisgrid.FacetGrid at 0x7f7ec53a44d0>



What does this mean now?

yet, we see that women paid more money for their Titanic-Adventure. There were much more Women wchich survived which were older than 10 years old. The passengers of first class had a much higher rate of survival. People in third class had the least survival chance. The third class was dominated by men in the age of around 25. People in first class below the age of 60 had the highest chance of survival. Especially children had a really high chance of suvival just like women.