



Chapter 7 - ex 4: Titanic

Cho dữ liệu titanic có sẵn trong seaborn library. Hãy vẽ những biểu đồ sau:

1. Vẽ stripplot thể hiện sự phân bố của fare theo class
2. Vẽ violinplot thể hiện sự phân bố của fare theo class
3. Vẽ countplot đếm số lượng alive/not alive theo từng class
4. Vẽ pointplot thể hiện khả năng sống sót 'survived' theo class
5. Vẽ factorplot (phiên bản mới là catplot) dạng bar thể hiện survived của từng sex, phân loại theo class
6. Vẽ correlation matrix (heatmap) của titanic

```
In [1]: import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
```

```
In [2]: # Load data
titanic = sns.load_dataset("titanic")
titanic.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
survived      891 non-null int64
pclass        891 non-null int64
sex           891 non-null object
age           714 non-null float64
sibsp         891 non-null int64
parch         891 non-null int64
fare          891 non-null float64
embarked      889 non-null object
class         891 non-null category
who           891 non-null object
adult_male    891 non-null bool
deck          203 non-null category
embark_town   889 non-null object
alive         891 non-null object
alone         891 non-null bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.6+ KB
```

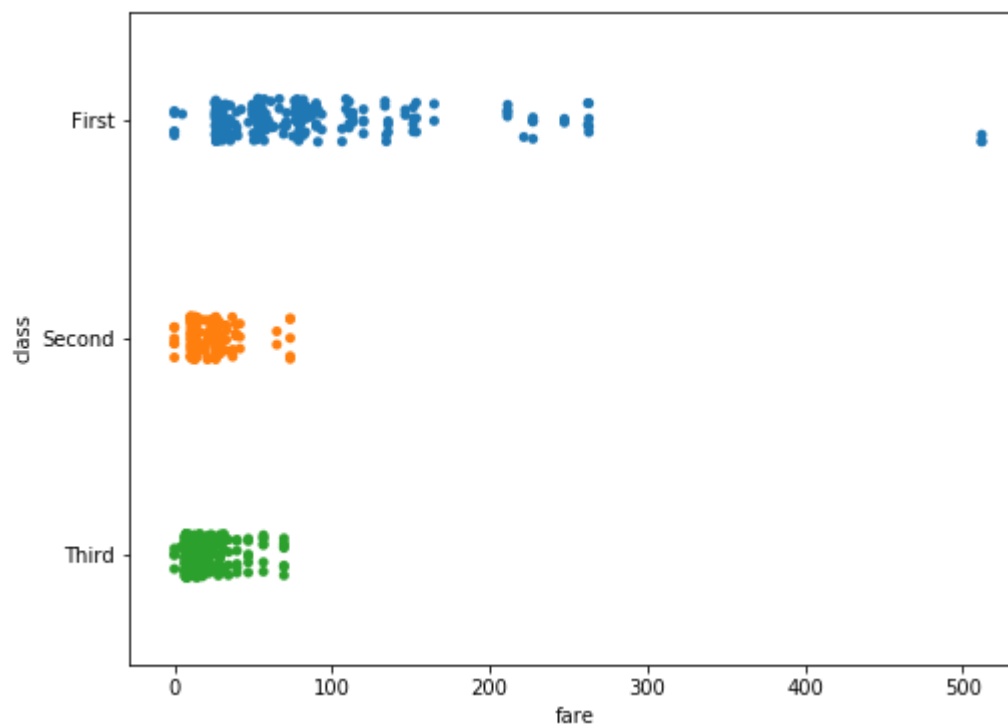
In [4]: `titanic.head()`

Out[4]:

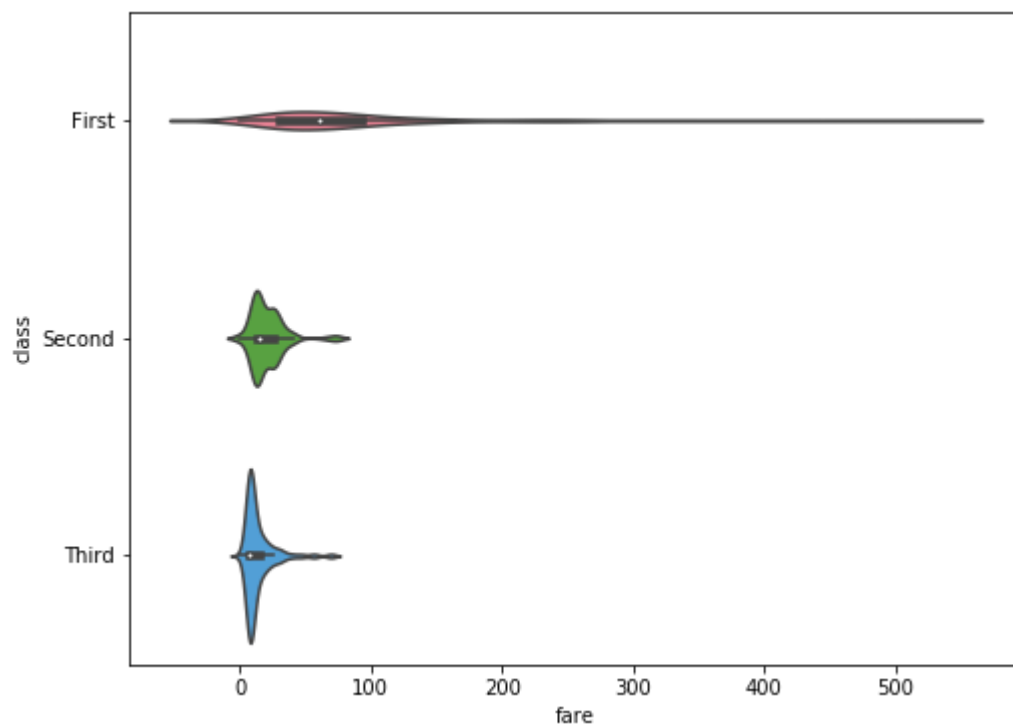
	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	de
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	N
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	N
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	N

In [6]:

```
# Create the stripplot
plt.figure(figsize=(8,6))
sns.stripplot(data=titanic,
              x='fare',
              y='class',
              jitter=True)
plt.show()
```

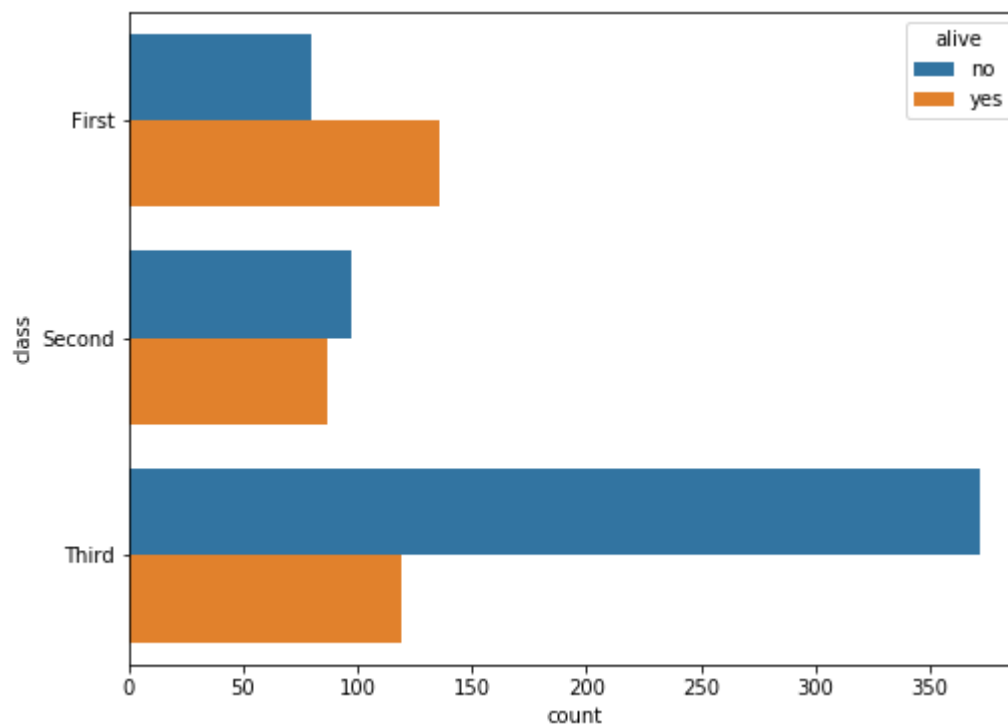


```
In [8]: # Create a violinplot with the husl palette
plt.figure(figsize=(8,6))
sns.violinplot(data=titanic,
               x='fare',
               y='class',
               palette='husl')
plt.show()
plt.clf()
```



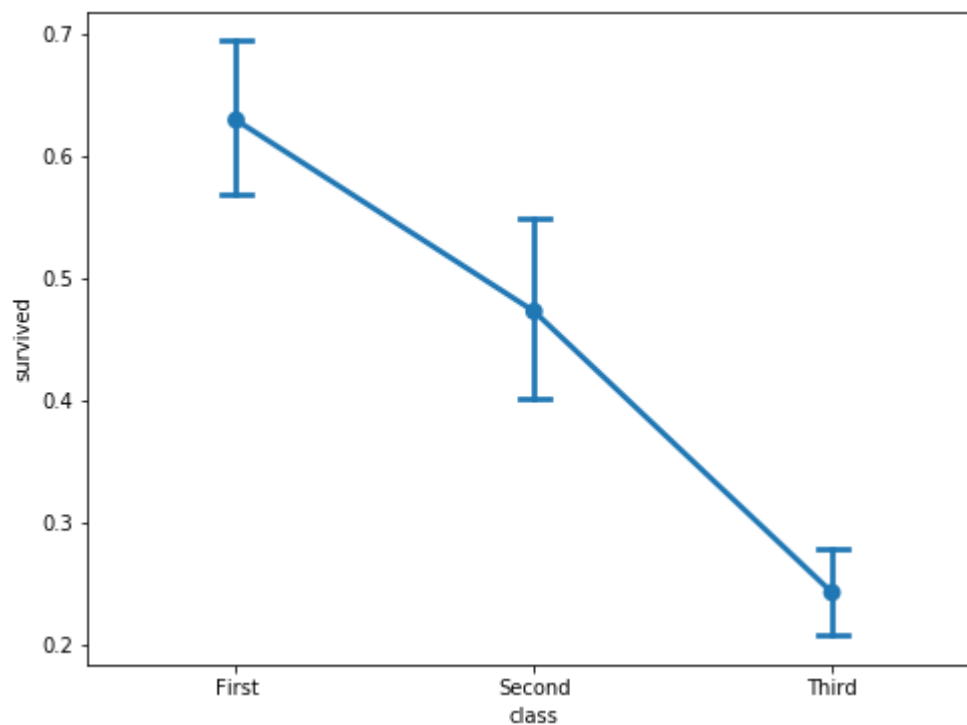
<Figure size 432x288 with 0 Axes>

```
In [13]: # Show a countplot with the number of models used
# with each region a different color
plt.figure(figsize=(8,6))
sns.countplot(data=titanic,
              y="class",
              hue="alive")
plt.show()
plt.clf()
```



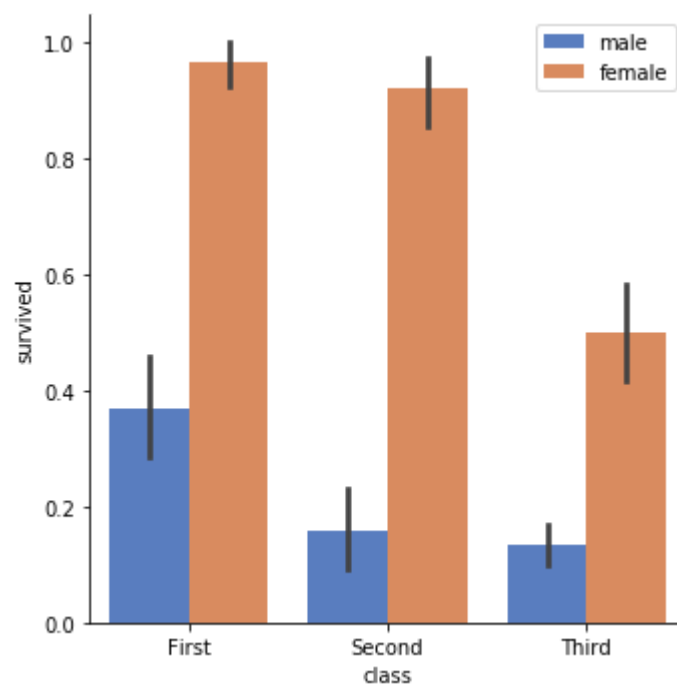
<Figure size 432x288 with 0 Axes>

```
In [16]: # Create a pointplot and include the capsize
# in order to show bars on the confidence interval
plt.figure(figsize=(8,6))
sns.pointplot(data=titanic,
              y='survived',
              x='class',
              capsize=.1)
plt.show()
plt.clf()
```

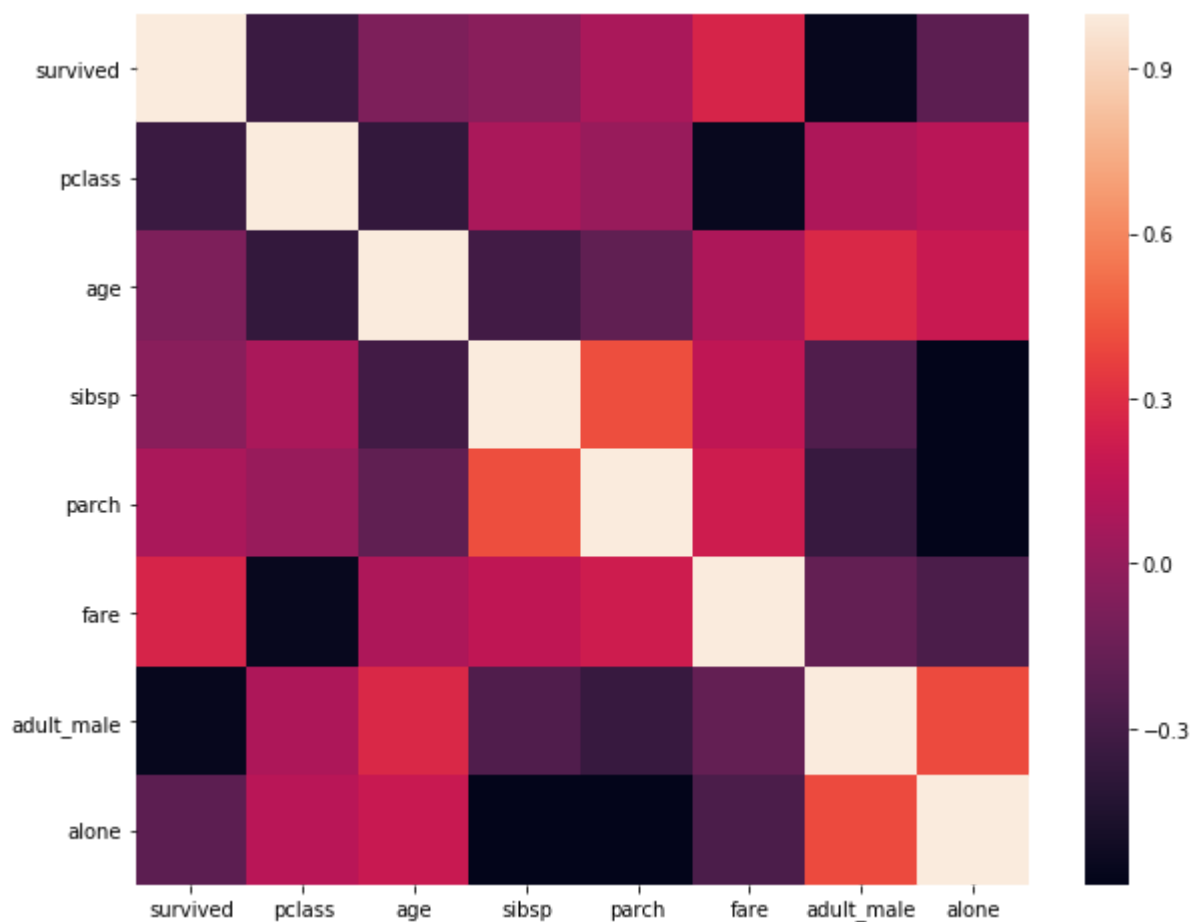


<Figure size 432x288 with 0 Axes>

```
In [3]: # Set up a factorplot  
g = sns.catplot("class", "survived", "sex", data=titanic, kind="bar", palette="mut  
plt.legend()  
# Show plot  
plt.show()
```



```
In [15]: # correlation matrix - heat map  
plt.figure(figsize=(10,8))  
sns.heatmap(titanic.corr())  
plt.show()
```



```
In [ ]:
```