```
What is the probability that a child who is in third class and is 10 years old or younger
survives?
How much did people pay to be on the ship (average)?
Calculate the expectation of fare conditioned on passenger-class (lowest and highest paid).
```

Richa Patel

Titanic Part 1

In [133]:
```
import pandas as pd
import numpy as np


import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
sns.set() # setting seaborn default for plots
sns.set(style="white", color_codes=True)

# read file

read = pd.read_csv('titanic.csv')
print(read)

print("\nInfo\n")
read.info()

print("Titanic Shape:",read.shape)

#We can see that there are 887 rows and 8 columns in our dataset.

read.head(10)
```

```
titanic.head(10)
```

```
     Survived  Pclass                                                Name  \
0           0       3                             Mr. Owen Harris Braund
1           1       1  Mrs. John Bradley (Florence Briggs Thayer) Cum...
2           1       3                         Miss. Laina Heikkinen
3           1       1       Mrs. Jacques Heath (Lily May Peel) Futrelle
4           0       3                       Mr. William Henry Allen
..        ...     ...                                               ...
882         0       2                         Rev. Juozas Montvila
883         1       1                  Miss. Margaret Edith Graham
884         0       3                Miss. Catherine Helen Johnston
885         1       1                       Mr. Karl Howell Behr
886         0       3                         Mr. Patrick Dooley

        Sex   Age  Siblings/Spouses Aboard  Parents/Children Aboard     Fare
0      male  22.0                        1                        0   7.2500
1    female  38.0                        1                        0  71.2833
2    female  26.0                        0                        0   7.9250
3    female  35.0                        1                        0  53.1000
4      male  35.0                        0                        0   8.0500
..      ...   ...                      ...                      ...      ...
882    male  27.0                        0                        0  13.0000
883  female  19.0                        0                        0  30.0000
884  female   7.0                        1                        2  23.4500
885    male  26.0                        0                        0  30.0000
886    male  32.0                        0                        0   7.7500

[887 rows x 8 columns]

Info

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 887 entries, 0 to 886
Data columns (total 8 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
```

```
 0   Survived                 887 non-null    int64
 1   Pclass                   887 non-null    int64
 2   Name                     887 non-null    object
 3   Sex                      887 non-null    object
 4   Age                      887 non-null    float64
 5   Siblings/Spouses Aboard  887 non-null    int64
 6   Parents/Children Aboard  887 non-null    int64
 7   Fare                     887 non-null    float64
dtypes: float64(2), int64(4), object(2)
memory usage: 55.6+ KB
Titanic Shape: (887, 8)
```

Out[133]:

	Survived	Pclass	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
0	0	3	Mr. Owen Harris Braund	male	22.0	1	0	7.2500
1	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cum...	female	38.0	1	0	71.2833
2	1	3	Miss. Laina Heikkinen	female	26.0	0	0	7.9250
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35.0	1	0	53.1000
4	0	3	Mr. William Henry Allen	male	35.0	0	0	8.0500
5	0	3	Mr. James Moran	male	27.0	0	0	8.4583
6	0	1	Mr. Timothy J McCarthy	male	54.0	0	0	51.8625
7	0	3	Master. Gosta Leonard Palsson	male	2.0	3	1	21.0750
8	1	3	Mrs. Oscar W (Elisabeth Vilhelmina Berg) Johnson	female	27.0	0	2	11.1333
9	1	2	Mrs. Nicholas (Adele Achem) Nasser	female	14.0	1	0	30.0708

In [134]:
```
read.describe()

read["Fare"]
read.head()
```

Out[134]:

	Survived	Pclass	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
0	0	3	Mr. Owen Harris Braund	male	22.0	1	0	7.2500
1	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cum...	female	38.0	1	0	71.2833
2	1	3	Miss. Laina Heikkinen	female	26.0	0	0	7.9250
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35.0	1	0	53.1000
4	0	3	Mr. William Henry Allen	male	35.0	0	0	8.0500

Data Visualizations

What is the probability that a child who is in third class and is 10 years old or younger survives?

In [135]:
```
def bar_chart(feature):
    pclass = read[read['Pclass']==3][feature].value_counts()
    age = read[read['Age']>=10][feature].value_counts()
    df = pd.DataFrame([pclass,age])
    df.index = ['Pclass','Age']
    df.plot(kind='bar',stacked=True, figsize=(10,5))
```

```
bar_chart('Sex')
print("pclass :\n",read[read['Pclass']==3]['Sex'].value_counts())


print("age:\n",read[read['Age']>=10]['Sex'].value_counts())


x = read.groupby('Age')
x.head()


#child = read.groupby('Age').sum()['Pclass' == 3]
#child2 = read.groupby('Age' > 10).sum()['Pclass']
#probability = child/child2
```

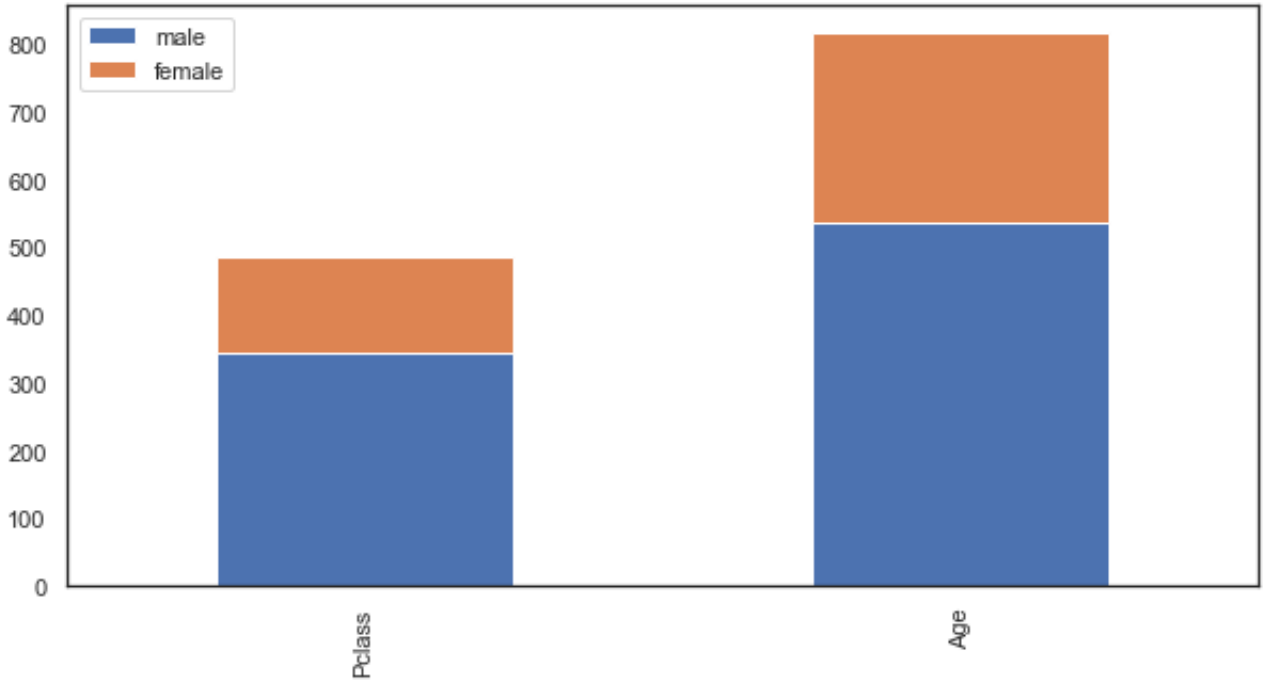
```
pclass :
 male      343
female    144
Name: Sex, dtype: int64
age:
 male      537
female    279
Name: Sex, dtype: int64
```

Out[135]:

	Survived	Pclass	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
0	0	3	Mr. Owen Harris Braund	male	22.0	1	0	7.2500
1	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cum...	female	38.0	1	0	71.2833
2	1	3	Miss. Laina Heikkinen	female	26.0	0	0	7.9250
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35.0	1	0	53.1000

4	0	3	Mr. William Henry Allen	male	35.0	0	0	8.0500
...	...	...	...	...	...	...	...	...
839	0	3	Mr. Peter L Lemberopolous	male	34.5	0	0	6.4375
847	0	3	Mr. Johan Svensson	male	74.0	0	0	7.7750
871	1	3	Miss. Adele Kiamie Najib	female	15.0	0	0	7.2250
875	1	1	Mrs. Thomas Jr (Lily Alexenia Wilson) Potter	female	56.0	0	1	83.1583
884	0	3	Miss. Catherine Helen Johnston	female	7.0	1	2	23.4500

326 rows × 8 columns



As we can see who's age is 10 or Grater is Male: 537 and female: 279

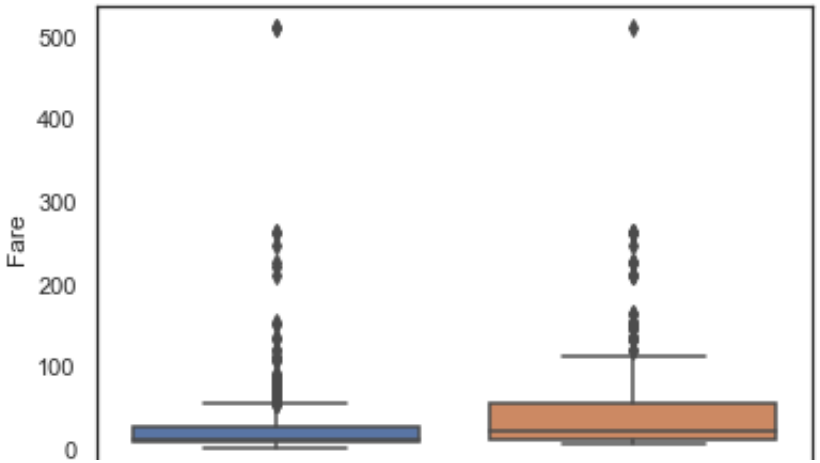
who is in 3rd class is male:343 and female: 144

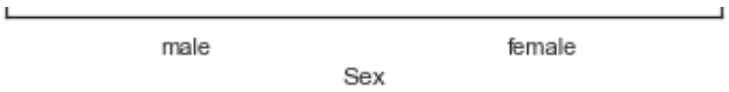
How much did people pay to be on the ship (average)?

In [136]:
```
avg= read['Fare'].mean()
```

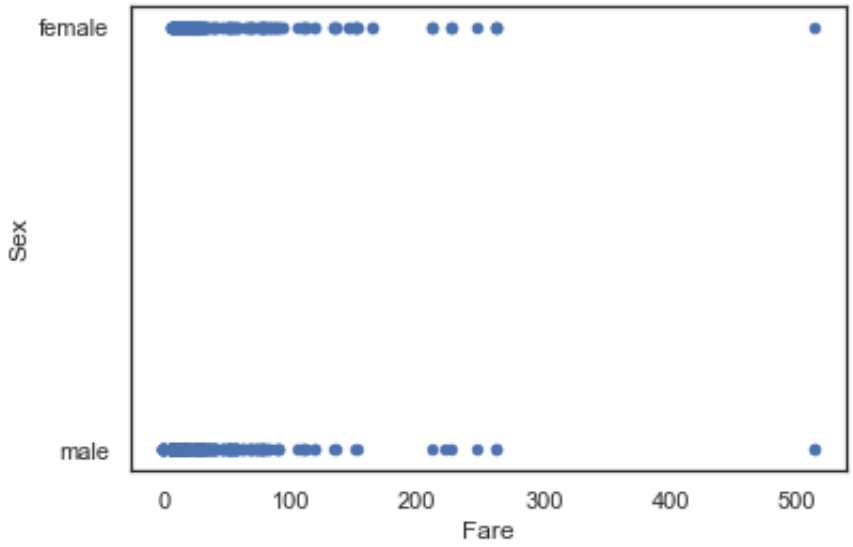
As We can see Peloe pay ava 32.305

In [137]:
```
sns.boxplot(x="Sex", y="Fare", data=read)
plt.show()

read.plot(kind="scatter",x="Fare" , y="Sex")
plt.show()
```





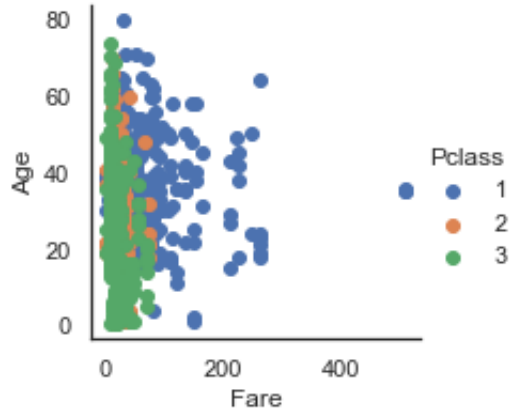

```
*c* argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value
-mapping will have precedence in case its length matches with *x* & *y*.  Please use the *color
* keyword-argument or provide a 2-D array with a single row if you intend to specify the same R
GB or RGBA value for all points.
```

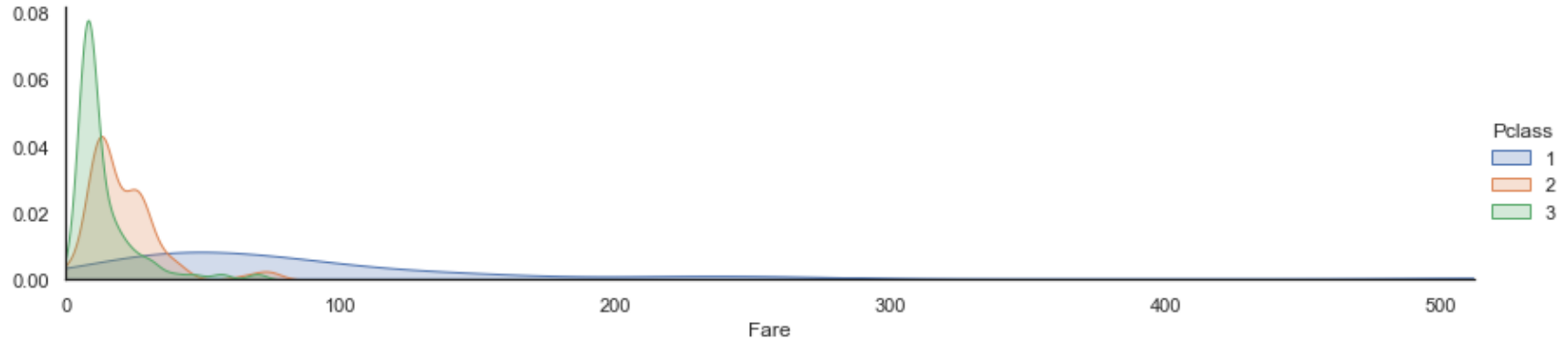


Calculate the expectation of fare conditioned on passenger-class (lowest and highest paid).

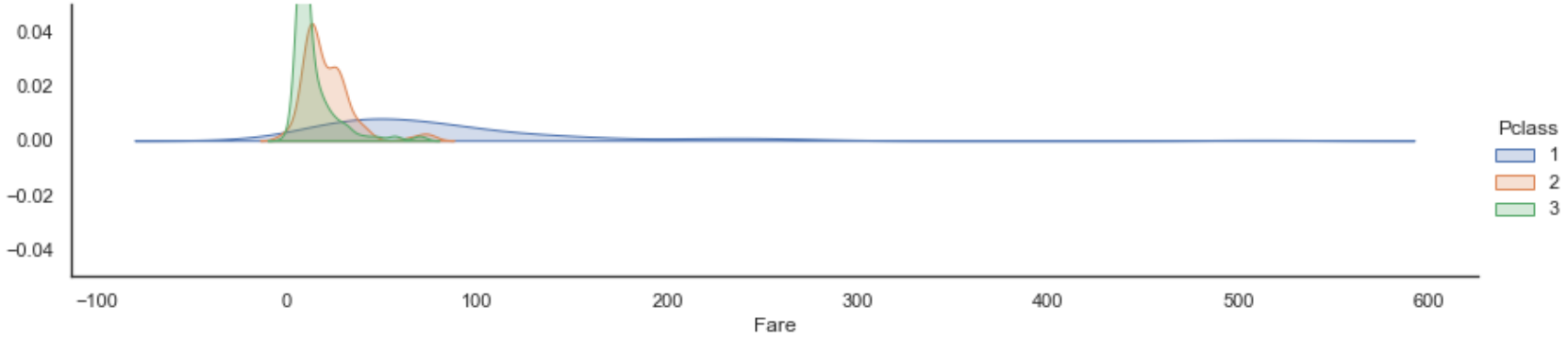
In [138]:
```
read["Fare"].fillna(read.groupby("Pclass")["Fare"].transform("median"), inplace=True)
read.head(5)


# scatter plot of male and femal  by fare
sns.FacetGrid(read, hue ="Pclass").map(plt.scatter, 'Fare','Age').add_legend()
plt.show()
```

```
plt.show()

facet = sns.FacetGrid(read, hue="Pclass",aspect=4 )
facet.map(sns.kdeplot, 'Fare', shade = True)
facet.set(xlim = (0, read['Fare'].max()))
facet.add_legend()
plt.show()

facet = sns.FacetGrid(read, hue="Pclass",aspect=4 )
facet.map(sns.kdeplot, 'Fare', shade = True)
facet.set(ylim = (0, read['Fare'].min()))
facet.add_legend()
plt.show()
```



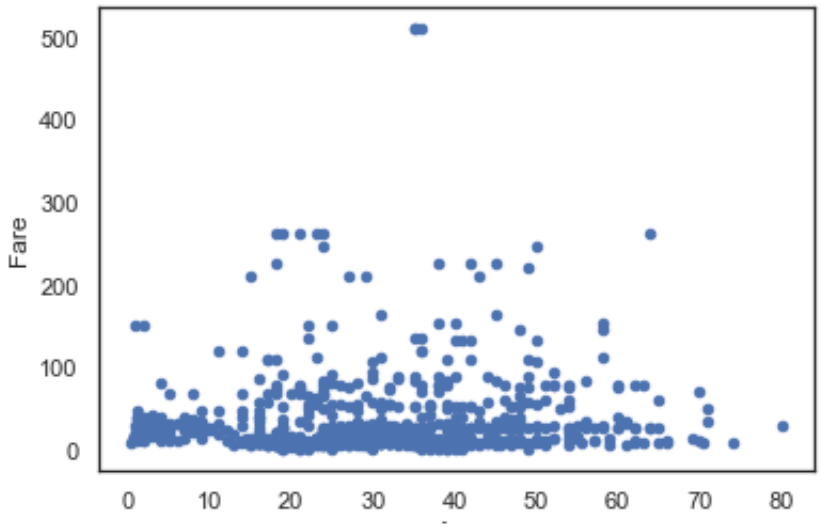


```
/Users/richapatel/opt/anaconda3/lib/python3.8/site-packages/seaborn/axisgrid.py:49: UserWarning
: Attempting to set identical bottom == top == 0 results in singular transformations; automatic
ally expanding.
  ax.set(**kwargs)
```



In [139]:
```
read.plot(kind="scatter",x="Age" , y="Fare")
plt.show()



# scatter plot of male and femal  by fare
sns.FacetGrid(read, hue ="Fare", height =5).map(plt.scatter, 'Sex','Fare')
plt.show()
```

```
*c* argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value
-mapping will have precedence in case its length matches with *x* & *y*.  Please use the *color
* keyword-argument or provide a 2-D array with a single row if you intend to specify the same R
GB or RGBA value for all points.
```



In [ ]:

In [ ]: