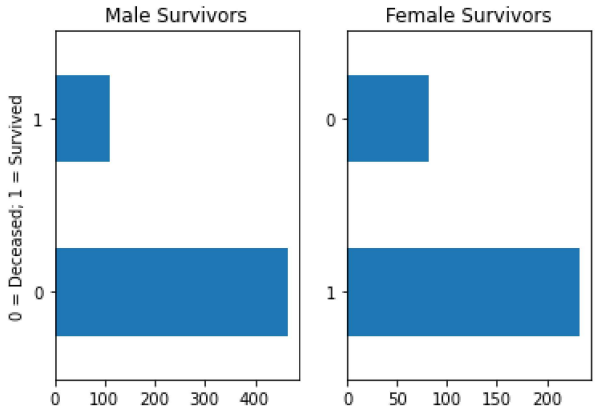
In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

data=pd.read_csv("C:/Learning/MS Degree/MIS542/titanic.csv", delimiter=',',engine='python')
data=pd.DataFrame(data)
print

# currConvert =pd.DataFrame([[1, 1.00, 'GBP'],[2, 1.28, 'USD']], columns=['id','value', 'currency'])
data['currency']='GBP'
data['Fare in USD']= data["Fare in British Pounds"] * 1.28

avgCost = data["Fare in USD"].mean()
avgCostFirstClass = data.loc[data['Passenger Class'] == 1, "Fare in USD"].mean()
hasSibling = data['Name'].loc[(data['Siblings/Spouses Aboard'] > 1) & (data['Age'] > 19)].count()
hasSiblingAlive = data['Name'].loc[(data['Siblings/Spouses Aboard'] > 1) & (data['Age'] > 19) & (data['Survived
hasSiblingDeceased = data['Name'].loc[(data['Siblings/Spouses Aboard'] > 1) & (data['Age'] > 19) & (data['Survi
meanAgeDeceased = data.loc[data['Survived'] == 0, "Age"].mean()
meanAgeSurvived = data.loc[data['Survived'] == 1, "Age"].mean()


print('Average Cost')
print(avgCost)
print('Average First Class Cost')
print(avgCostFirstClass)
print('Count Of Passengers 20 And Over With Siblings')
print(hasSibling)
print('Additional Question 1: Show Count Of Surviving Passengers 20 And Over With Siblings')
print(hasSiblingAlive)
print('Additional Question 2: Show Count Of Deceased Passengers 20 And Over With Siblings')
print(hasSiblingDeceased)
print('Average age of Deceased')
print(meanAgeDeceased)
print('Average age of Surviving')
print(meanAgeSurvived)

data['Gender'] = 0

sex = np.array(data['Sex'])
data.loc[data['Sex'] == 'male', 'Gender'] = 0
data.loc[data['Sex'] == 'female', 'Gender'] = 1
# male = data['Name'].loc[(data['Sex'] == 'male')].count()
# female = data['Name'].loc[(data['Sex'] == 'female')].count()
```

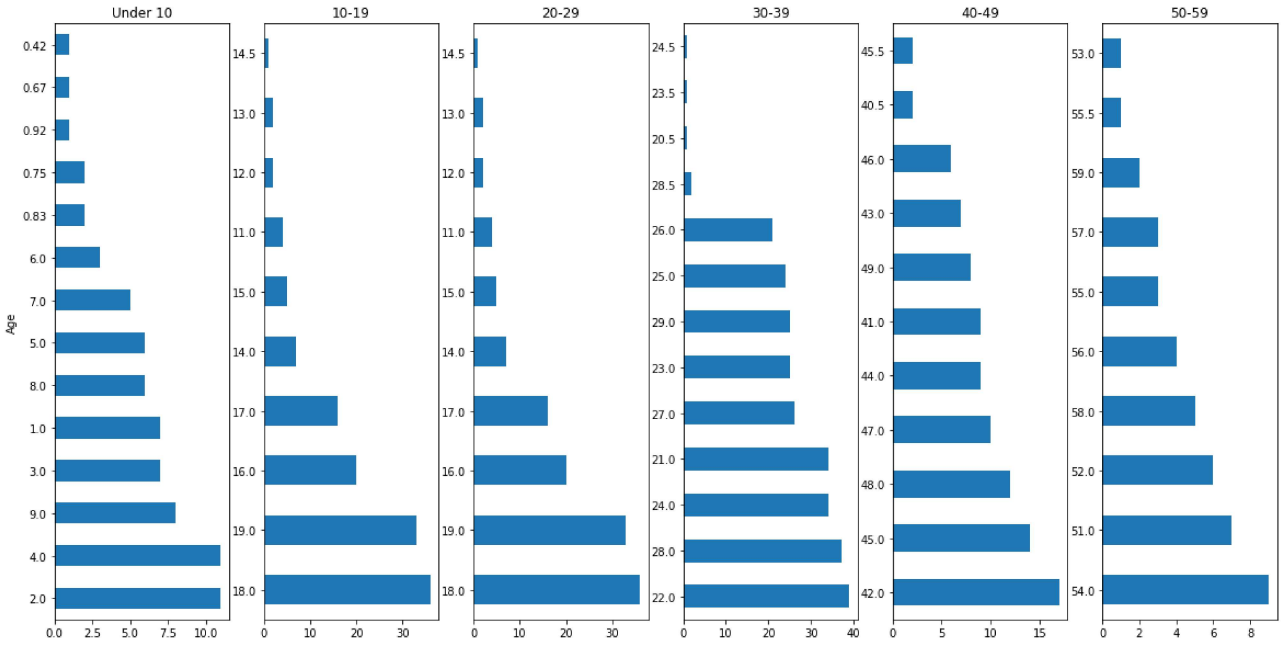
```
Average Cost
41.35093783089063
Average First Class Cost
107.71800000000006
Count Of Passengers 20 And Over With Siblings
24
Additional Question 1: Show Count Of Surviving Passengers 20 And Over With Siblings
11
Additional Question 2: Show Count Of Deceased Passengers 20 And Over With Siblings
13
Average age of Deceased
30.13853211009174
Average age of Surviving
28.408391812865496
```

In [2]:
```
fig, axs = plt.subplots(1,2)
data[data.Sex=='male'].Survived.value_counts().plot(kind='barh',ax=axs[0], title='Male Survivors', ylabel="0 = D
data[data.Sex=='female'].Survived.value_counts().plot(kind='barh',ax=axs[1], title='Female Survivors')
```

Out[2]:
```
<AxesSubplot:title={'center':'Female Survivors'}>
```

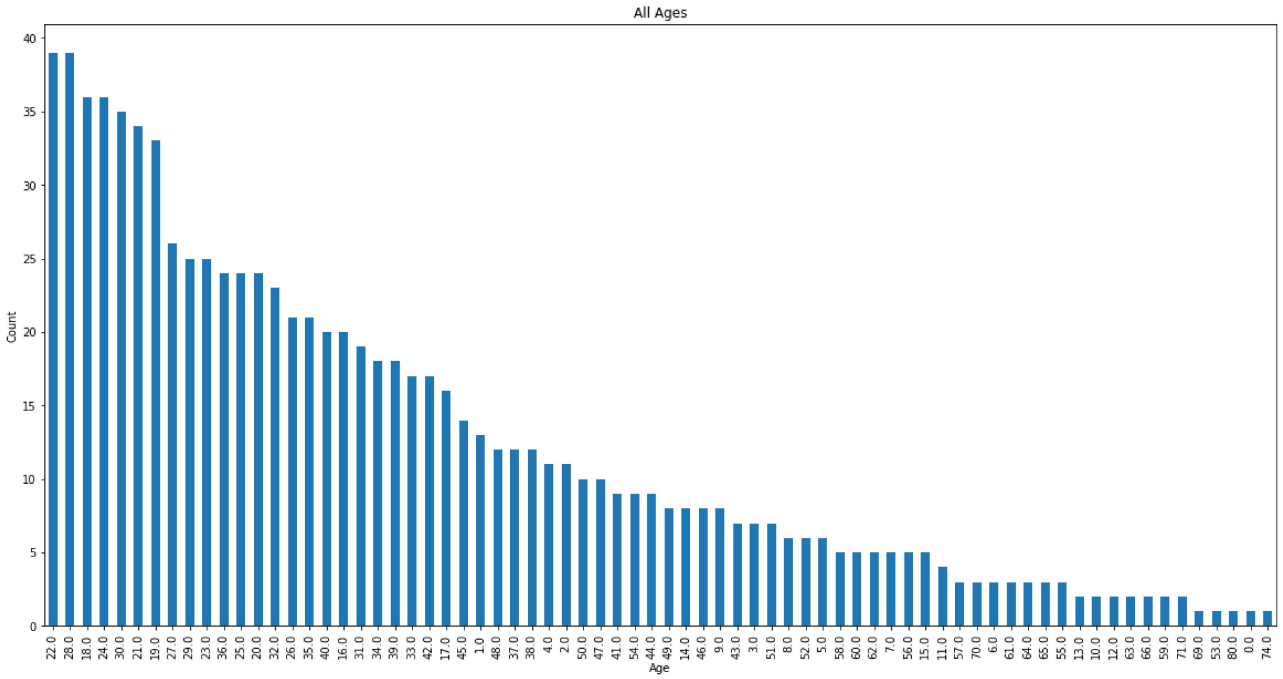


In [3]:
```
fig, axs = plt.subplots(1,6)
data[(data.Age<10)].Age.value_counts().plot(kind='barh',ax=axs[0], title='Under 10',figsize=(20,10), ylabel='Age
data[(data.Age>10) & (data.Age<20)].Age.value_counts().plot(kind='barh',ax=axs[1], title='10-19')
data[(data.Age>10) & (data.Age<20)].Age.value_counts().plot(kind='barh',ax=axs[2], title='20-29')
data[(data.Age>20) & (data.Age<30)].Age.value_counts().plot(kind='barh',ax=axs[3], title='30-39')
data[(data.Age>40) & (data.Age<50)].Age.value_counts().plot(kind='barh',ax=axs[4], title='40-49')
data[(data.Age>50) & (data.Age<60)].Age.value_counts().plot(kind='barh',ax=axs[5], title='50-59')
plt.show(1,6)
# data[data.Age>2].Age.value_counts().plot(kind='bar',ax=axs[1])
```

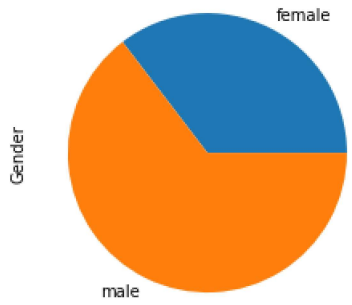


In [4]:
```
data[(data.Age>0)].Age.round().value_counts().plot(kind='bar', title='All Ages', figsize=(20,10), xlabel = 'Age
```

Out[4]:
```
<AxesSubplot:title={'center':'All Ages'}, xlabel='Age', ylabel='Count'>
```

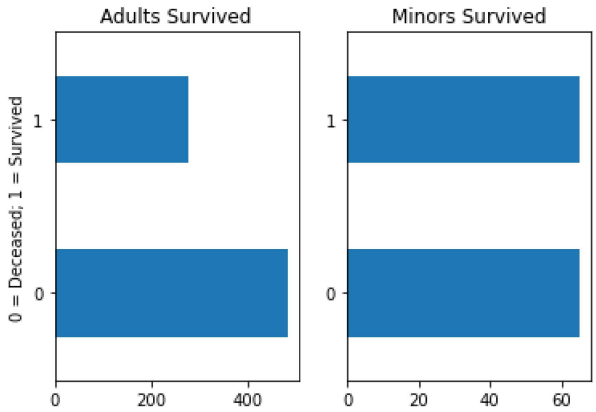


In [5]:
```
data.loc[data['Sex'] == 'male', 'Gender'] = 1
data.loc[data['Sex'] == 'female', 'Gender'] = 1
data.groupby('Sex')['Gender'].sum().plot(kind='pie')
plt.show()
data.loc[data['Sex'] == 'male', 'Gender'] = 0
data.loc[data['Sex'] == 'female', 'Gender'] = 1
```



In [6]:
```
fig, axs = plt.subplots(1,2)
data[data.Age>=18].Survived.value_counts().plot(kind='barh',ax=axs[0], title='Adults Survived', ylabel="0 = Dece
data[data.Age<=17.99].Survived.value_counts().plot(kind='barh',ax=axs[1], title='Minors Survived')
```

Out[6]:
```
<AxesSubplot:title={'center':'Minors Survived'}>
```

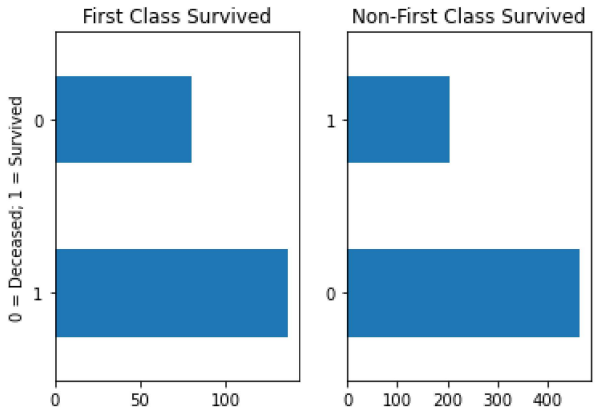


In [7]:
```
data.columns = data.columns.str.replace(' ','_') # Easiest way to call spaces in Pandas seems to be to replace t
data.columns = data.columns.str.replace('/','_') # Easiest way to call spaces in Pandas seems to be to replace t
```

```
fig, axs = plt.subplots(1,2)
data[data.Passenger_Class==1].Survived.value_counts().plot(kind='barh',ax=axs[0], title='First Class Survived',
data[data.Passenger_Class!=1].Survived.value_counts().plot(kind='barh',ax=axs[1], title='Non-First Class Survive
```

Out[7]:

```
<AxesSubplot:title={'center':'Non-First Class Survived'}>
```

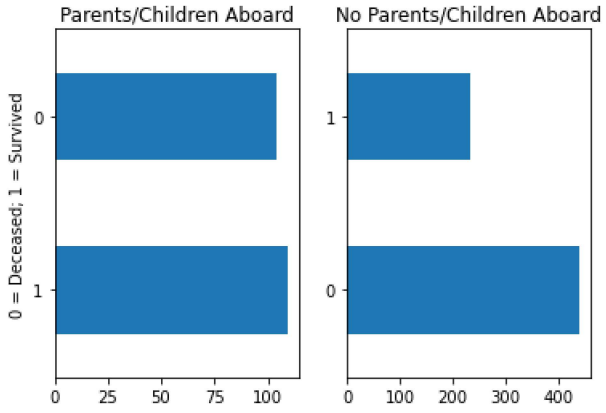


In [8]:

```
fig, axs = plt.subplots(1,2)
data[data.Parents_Children_Aboard>0].Survived.value_counts().plot(kind='barh',ax=axs[0], title='Parents/Children
data[data.Parents_Children_Aboard==0].Survived.value_counts().plot(kind='barh',ax=axs[1], title='No Parents/Chi
```

Out[8]:

```
<AxesSubplot:title={'center':'No Parents/Children Aboard'}>
```



In [9]:

```
data
```

Out[9]:

	Survived	Passenger_Class	Name	Sex	Age	Siblings_Spouses_Aboard	Parents_Children_Aboard	Fare_in_British_Pounds	c
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	
...	...	...	...	...	...	...	...	...	
882	0	2	Rev. Juozas Montvila	male	27.0	0	0	13.0000	
883	1	1	Miss. Margaret Edith Graham	female	19.0	0	0	30.0000	
884	0	3	Miss. Catherine Helen Johnston	female	7.0	1	2	23.4500	
885	1	1	Mr. Karl Howell Behr	male	26.0	0	0	30.0000	
886	0	3	Mr. Patrick Dooley	male	32.0	0	0	7.7500	

887 rows × 11 columns

In [ ]: