import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

titanic\_df=pd.read\_csv("titanic\_data.csv")

titanic\_df.head()

titanic\_df.info()

titanic\_df.describe()

"""

######### Homework\_1(a) ################

**(a) Plot a histogram detailing the age distribution of the dataset. What do you conclude from this histogram?**

"""

#age = titanic\_df.iloc[:, 3].values

age = titanic\_df.loc[:, 'Age'].values

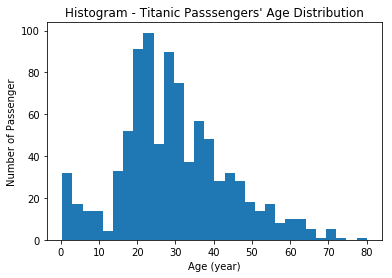
plt.hist(age, bins = 30)

plt.title("Histogram - Titanic Passsengers' Age Distribution")

plt.xlabel("Age (year)")

plt.ylabel("Number of Passenger")

plt.show()



"""

**Conclusion:**

The Histogram plot refers that most of the titanic passengers were between the age 20-25 and 30-35 years. About 350 were in these age group. There we about 78 children; and only about 10 were 65 years or older. Only about 5 children were between 10-15 years old and 2-3 passengers were 72-80 years old.

"""

"""

######### Homework\_1(b) ################

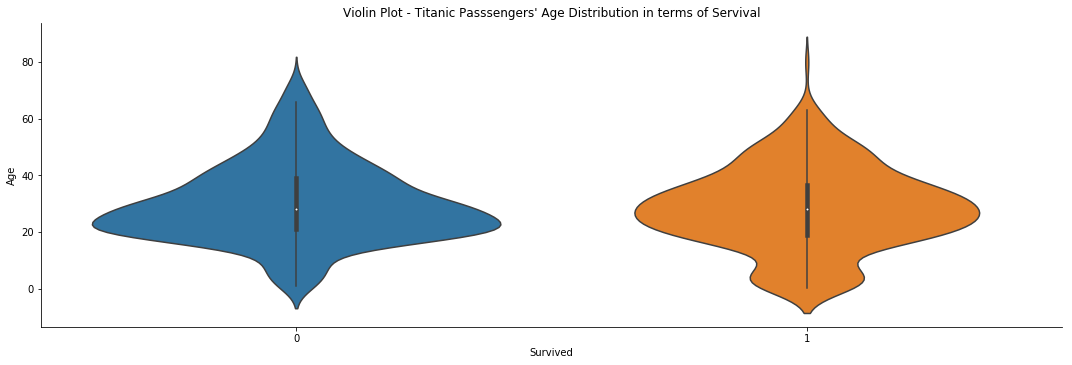
**(b) Show a violin plot detailing the age distribution, separated by survivors and deceased. What do you conclude from this plot?**

"""

sns.factorplot(data=titanic\_df,x='Survived',y='Age',kind="violin",size=5,aspect=3)

#plt.yticks([0,1], ['Decceased (0)' ,'Survived (1)'])

plt.title("Violin Plot - Titanic Passsengers' Age Distribution in terms of Servival",fontsize=12);



"""

**Conclusion:**

The Violin plots refer that most of the deceased passengers were between 20 and 35 years old.

Althouh most of the survived passengers were approximately within the above age groups there were

a few survilals between 10-15 years old; and above 60 years old.

"""

"""

######### Homework\_1(c) ################

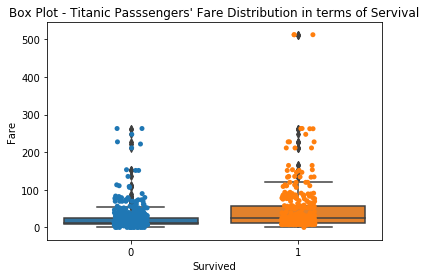
**(c) Show a box plot summarizing the fare distribution, separated by survivors and deceased. What do you conclude from this plot?**

"""

sns.boxplot(x="Survived", y="Fare", data=titanic\_df)

sns.stripplot(x="Survived", y="Fare", data=titanic\_df, jitter=True, edgecolor="gray")

plt.title("Box Plot - Titanic Passsengers' Fare Distribution in terms of Servival",fontsize=12);



"""

**Conclusion:**

The box plot refers that among the deceased passengers most of them paid around $30- $50 fare;

a few of them paid between $210 - $280. On the other hand, among the deceased passengers

most of them paid slightly higher than the deceased group of passengers. At least some of them

them paid $500 or more fare.

"""

"""

######### Homework\_1(d) ################

**(d) Show a scatter plot of gender vs. survival. What do you conclude from this plot?**

"""

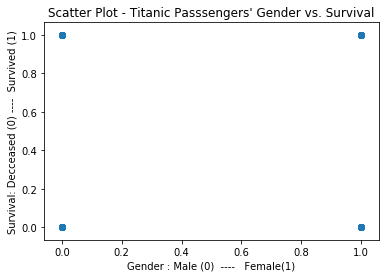
plt.scatter(titanic\_df.Survived, titanic\_df.Sex, alpha=0.1) ## here the plot has to be transparent so we need to pic low alpha value

plt.title("Scatter Plot - Titanic Passsengers' Gender vs. Survival")

plt.xlabel("Gender : Male (0) ---- Female(1)")

plt.ylabel("Survival: Deceased (0) ---- Survived (1)")

plt.show()



"""

**Conclusion:**

The scatter plot does not explicitly refer any correlation between the titanic passengers'

gender vs survivlas. Both male and female passengers deceased and survived.

"""

"""

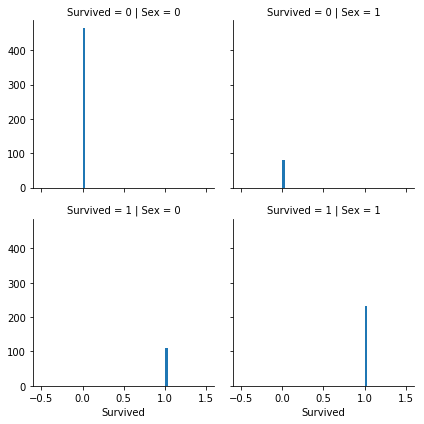
######### Homework\_1(e) ################

**(e) Plot a histogram showing the distribution of survivors as a function of gender. What do you conclude from this histogram?**

"""

surv\_sex\_hist = sns.FacetGrid(titanic\_df, row='Survived', col='Sex')

surv\_sex\_hist.map(plt.hist, 'Survived', bins=30)



"""

**Conclusion:**

A histogram plot showing the distribution of survivors as a function of gender infers that

among the titanic passengers, about 460 male and 80 female were deceased;

about 110 male and 240 female were survived.

"""

"""

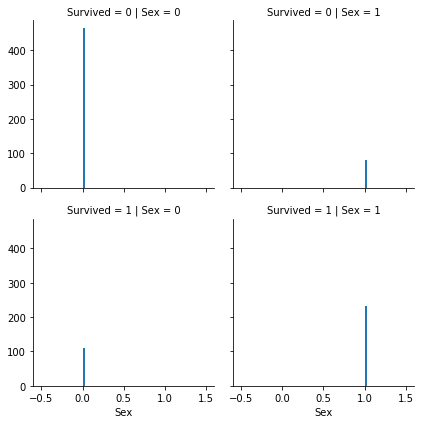
######### Homework\_1(f) ################

**(f) Plot a histogram showing the distribution of gender as a function of survival. What do you conclude from this histogram?**

"""

surv\_sex\_hist = sns.FacetGrid(titanic\_df, col='Sex', row='Survived')

surv\_sex\_hist.map(plt.hist, 'Sex', bins=30)



"""

**Conclusion:**

A histogram plot showing the distribution of survivors as a function of gender infers that

among the titanic passengers, about 460 male and 80 female were deceased;

about 110 male and 240 female were survived. Seems, the (e) and (f) depict identically.

"""