Question 2

2a.

Make bar plots of intent for each combination of race and sex. For each race, make a single plot with subplots for the two sexes.

In [2]:
```
#load necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [17]:
```
gundata = pd.read_csv("C:/Users/Mary Akinde/Documents/Modern Statistics/full_data
gundata.shape
```

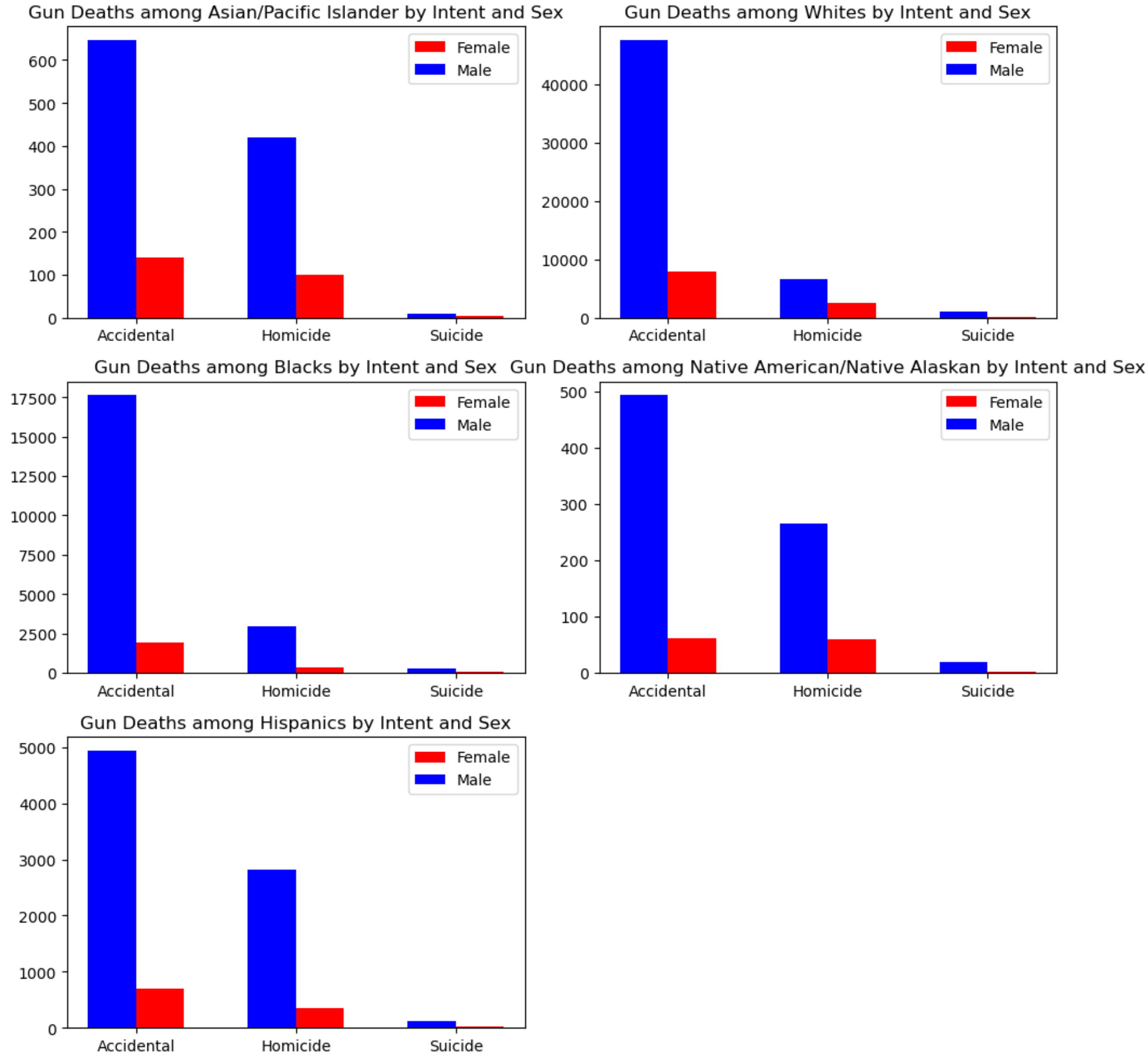
Out[17]:
```
(100798, 11)
```

In [18]:
```
#Each combination as race and sex excluding Undetermined
Asian_F =gundata[(gundata.iloc[:,7]=="Asian/Pacific Islander")&(gundata.iloc[:,5]=
Asian_M =gundata[(gundata.iloc[:,7]=="Asian/Pacific Islander")&(gundata.iloc[:,5]=
White_F =gundata[(gundata.iloc[:,7]=="White")&(gundata.iloc[:,5]=="F")&(gundata.il
White_M =gundata[(gundata.iloc[:,7]=="White")&(gundata.iloc[:,5]=="M")&(gundata.il
Black_F =gundata[(gundata.iloc[:,7]=="Black")&(gundata.iloc[:,5]=="F")&(gundata.il
Black_M =gundata[(gundata.iloc[:,7]=="Black")&(gundata.iloc[:,5]=="M")&(gundata.il
Native_F =gundata[(gundata.iloc[:,7]=="Native American/Native Alaskan")&(gundata.i
Native_M =gundata[(gundata.iloc[:,7]=="Native American/Native Alaskan")&(gundata.i
Hispanic_F =gundata[(gundata.iloc[:,7]=="Hispanic")&(gundata.iloc[:,5]=="F")&(gund
Hispanic_M =gundata[(gundata.iloc[:,7]=="Hispanic")&(gundata.iloc[:,5]=="M")&(gund

#Each race except intent of Undetermined
Asian =gundata[(gundata.iloc[:,7]=="Asian/Pacific Islander")&(gundata.iloc[:,3]!="U
White =gundata[(gundata.iloc[:,7]=="White")&(gundata.iloc[:,3]!="Undetermined")]
Black =gundata[(gundata.iloc[:,7]=="Black")&(gundata.iloc[:,3]!="Undetermined")]
Native =gundata[(gundata.iloc[:,7]=="Native American/Native Alaskan")&(gundata.iloc
Hispanic =gundata[(gundata.iloc[:,7]=="Hispanic")&(gundata.iloc[:,3]!="Undetermine

#get counts in each race by intent
all_counts_asians=Asian.intent.value_counts()
all_counts_white=White.intent.value_counts()
all_counts_black=Black.intent.value_counts()
all_counts_native=Native.intent.value_counts()
all_counts_hispanic=Hispanic.intent.value_counts()

#counts of each race by sex for intent
counts_Asian_F=Asian_F.intent.value_counts()
counts_Asian_M=Asian_M.intent.value_counts()
counts_White_F=White_F.intent.value_counts()
counts_White_M=White_M.intent.value_counts()
counts_Black_F=Black_F.intent.value_counts()
counts_Black_M=Black_M.intent.value_counts()
counts_Native_F=Native_F.intent.value_counts()
counts_Native_M=Native_M.intent.value_counts()
counts_Hispanic_F=Hispanic_F.intent.value_counts()
counts_Hispanic_M=Hispanic_M.intent.value_counts()

##reorder by correct order of intent
intents=['Accidental','Homicide','Suicide']
all_counts_asians=all_counts_asians.reindex(intents)
```

```
all_counts_white=all_counts_white.reindex(intents)
all_counts_black=all_counts_black.reindex(intents)
all_counts_native=all_counts_native.reindex(intents)
all_counts_hispanic=all_counts_hispanic.reindex(intents)
```

In [19]:

```
xpos=np.arange(len(all_counts_asians))  #evenly spaced x positions
xpos1=np.arange(len(all_counts_white))  #evenly spaced x positions
xpos2=np.arange(len(all_counts_black))  #evenly spaced x positions
xpos3=np.arange(len(all_counts_native))  #evenly spaced x positions
xpos4=np.arange(len(all_counts_hispanic))  #evenly spaced x positions

#plot
plt.figure(figsize=(10,10))
plt.subplot(321)
plt.bar(x=xpos+0.15,height=counts_Asian_F,label="Female",width=0.3,color="red")
plt.bar(x=xpos-0.15,height=counts_Asian_M,width=0.3,label="Male",color="blue")
plt.title("Gun Deaths among Asian/Pacific Islander by Intent and Sex")
plt.xticks(xpos,intents)
plt.legend()

plt.subplot(322)
plt.bar(x=xpos1+0.15,height=counts_White_F,label="Female",width=0.3,color="red")
plt.bar(x=xpos1-0.15,height=counts_White_M,width=0.3,label="Male",color="blue")
plt.title("Gun Deaths among Whites by Intent and Sex")
plt.xticks(xpos1,intents)
plt.legend()

plt.subplot(323)
plt.bar(x=xpos2+0.15,height=counts_Black_F,label="Female",width=0.3,color="red")
plt.bar(x=xpos2-0.15,height=counts_Black_M,width=0.3,label="Male",color="blue")
plt.title("Gun Deaths among Blacks by Intent and Sex")
plt.xticks(xpos2,intents)
plt.legend()

plt.subplot(324)
plt.bar(x=xpos3+0.15,height=counts_Native_F,label="Female",width=0.3,color="red")
plt.bar(x=xpos3-0.15,height=counts_Native_M,width=0.3,label="Male",color="blue")
plt.title("Gun Deaths among Native American/Native Alaskan by Intent and Sex")
plt.xticks(xpos3,intents)
plt.legend()

plt.subplot(325)
plt.bar(x=xpos4+0.15,height=counts_Hispanic_F,label="Female",width=0.3,color="red"
plt.bar(x=xpos4-0.15,height=counts_Hispanic_M,width=0.3,label="Male",color="blue")
plt.title("Gun Deaths among Hispanics by Intent and Sex")
plt.xticks(xpos4,intents)
plt.legend()

plt.tight_layout()
plt.show()
```



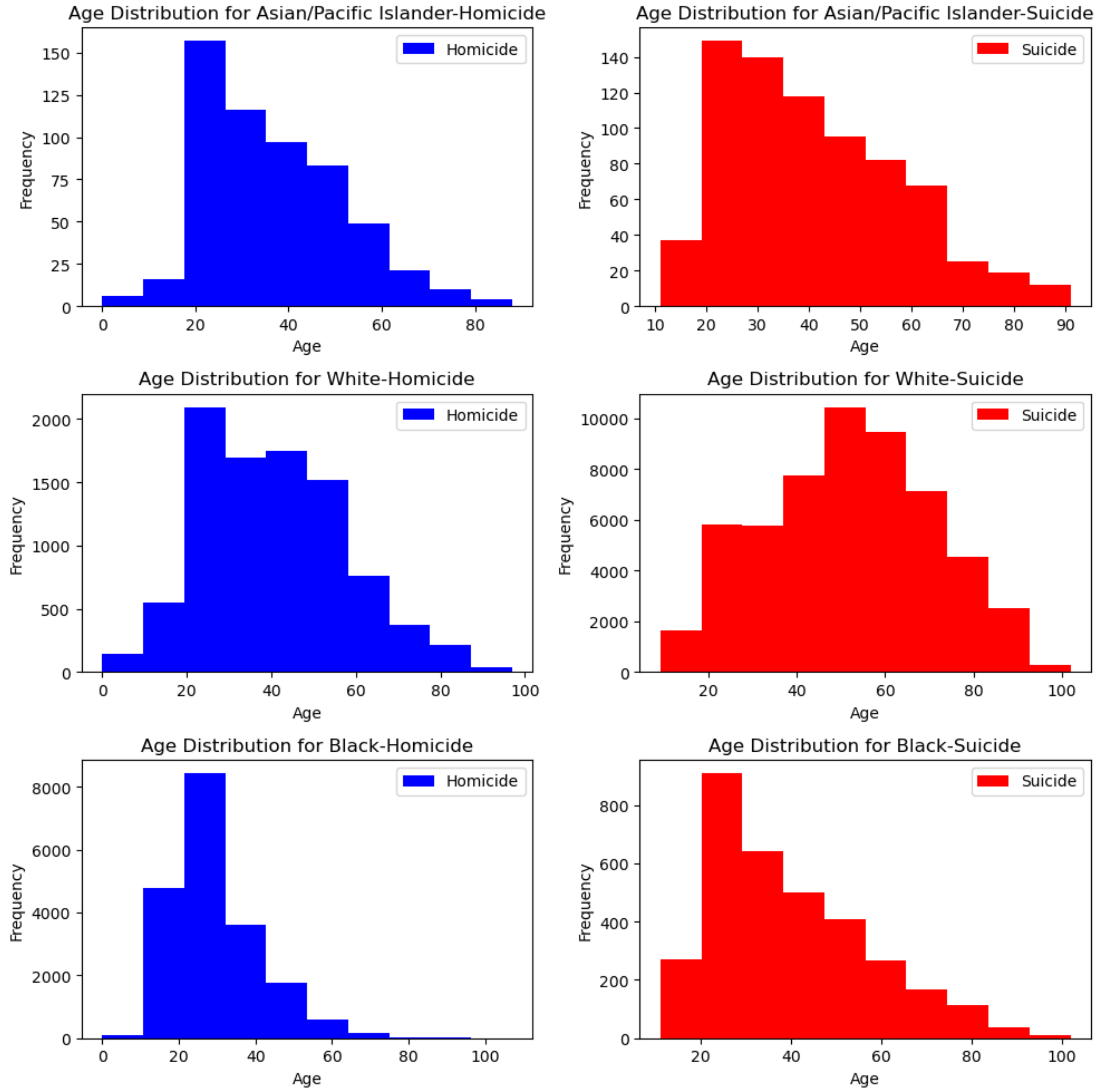
In general, males experience a higher number of gun-related fatalities compared to females. Accidental gun deaths are the most prevalent form of intent, followed by homicides and suicides. Among racial groups, white individuals have the highest incidence of gun deaths, followed by blacks, Hispanics, Asians, and Native Americans/Alaskans, respectively.

Question 2b

Make a plot of the age distribution for each homicides and suicides for your choice of three of the five races present in the data.

In [83]:
```
#Each combination as variables
Asian_H =gundata[(gundata.iloc[:,7]=="Asian/Pacific Islander")&(gundata.iloc[:,3]=
Asian_S =gundata[(gundata.iloc[:,7]=="Asian/Pacific Islander")&(gundata.iloc[:,3]=
White_H =gundata[(gundata.iloc[:,7]=="White")&(gundata.iloc[:,3]=="Homicide")]
White_S =gundata[(gundata.iloc[:,7]=="White")&(gundata.iloc[:,3]=="Suicide")]
Black_H =gundata[(gundata.iloc[:,7]=="Black")&(gundata.iloc[:,3]=="Homicide")]
Black_S =gundata[(gundata.iloc[:,7]=="Black")&(gundata.iloc[:,3]=="Suicide")]

plt.figure(figsize=(10, 10))

plt.subplot(321)
plt.hist(Asian_H.age, bins=10, alpha=1, label=Asian_H.intent, color = 'blue')
plt.title("Age Distribution for Asian/Pacific Islander-Homicide")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.legend()
```

```
plt.subplot(322)
plt.hist(Asian_S.age, bins=10, alpha=1, label=Asian_S.intent, color = 'red')
plt.title("Age Distribution for Asian/Pacific Islander-Suicide")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.legend()
plt.subplot(323)
plt.hist(White_H.age, bins=10, alpha=1, label=White_H.intent, color = 'blue')
plt.title("Age Distribution for White-Homicide")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.legend()
plt.subplot(324)
plt.hist(White_S.age, bins=10, alpha=1, label=White_S.intent, color = 'red')
plt.title("Age Distribution for White-Suicide")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.legend()
plt.subplot(325)
plt.hist(Black_H.age, bins=10, alpha=1, label=Black_H.intent, color = 'blue')
plt.title("Age Distribution for Black-Homicide")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.legend()
plt.subplot(326)
plt.hist(Black_S.age, bins=10, alpha=1, label=Black_S.intent, color = 'red')
plt.title("Age Distribution for Black-Suicide")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.legend()

plt.tight_layout()
plt.show()
```



Upon examining the provided plot for the age distribution of Asians/Pacific Islanders, we can observe that the distribution for both homicide and suicide intents is skewed to the right. This indicates that a larger proportion of gun-related deaths in this group occurs among individuals in younger age brackets(twenties).

Similarly, when considering the age distribution for homicide and suicide among Black individuals, we again see a right-skewed pattern. This suggests that a significant number of gun deaths among this group also occur among emerging adults individuals(twenties).

In contrast, the age distribution for both homicide and suicide among White individuals appears to be relatively symmetrical. This implies that gun-related fatalities in this group are spread more evenly across different age groups.

Question 2c

Interpretation 2c

Based on the plots above,the analysis reveals that males have a higher likelihood of experiencing gun-related fatalities compared to females. Accidental gun deaths are the most prevalent form of intent, followed by homicides and suicides. Among racial groups, white individuals have the highest rate of gun deaths, with blacks, Hispanics, Asians, and Native Americans/Alaskans following in decreasing order.

When examining the age distribution for Asians/Pacific Islanders, it is evident that gun-related deaths due to homicide and suicide are concentrated among younger individuals in their twenties, as indicated by the right-skewed distribution. Similarly, among Black individuals, gun deaths occur predominantly in the twenties, with a right-skewed age distribution for both homicide and suicide intents. On the other hand, the age distribution for White individuals shows a relatively symmetrical pattern, suggesting a more evenly distributed occurrence of gun-related fatalities across different age groups.

Question 3

In this problem, you will use the College Scorecard data to compare student debt and completion rates between public, private not-for-profit, and for-profit institutions. You should use Python for this problem.

In [20]:

```
##load data for college data
collegedata=pd.read_csv("C:/Users/Mary Akinde/Documents/Modern Statistics/MERGED20
collegedata.shape
```

```
C:\Users\Mary Akinde\AppData\Local\Temp\ipykernel_22236\3385437688.py:2: DtypeWarn
ing: Columns (6,9,1729,1743) have mixed types. Specify dtype option on import or s
et low_memory=False.
  collegedata=pd.read_csv("C:/Users/Mary Akinde/Documents/Modern Statistics/MERGED
2015_16_PP.csv",na_values=['NULL','PrivacySuppressed'])
```

Out[20]:

```
(7593, 1805)
```

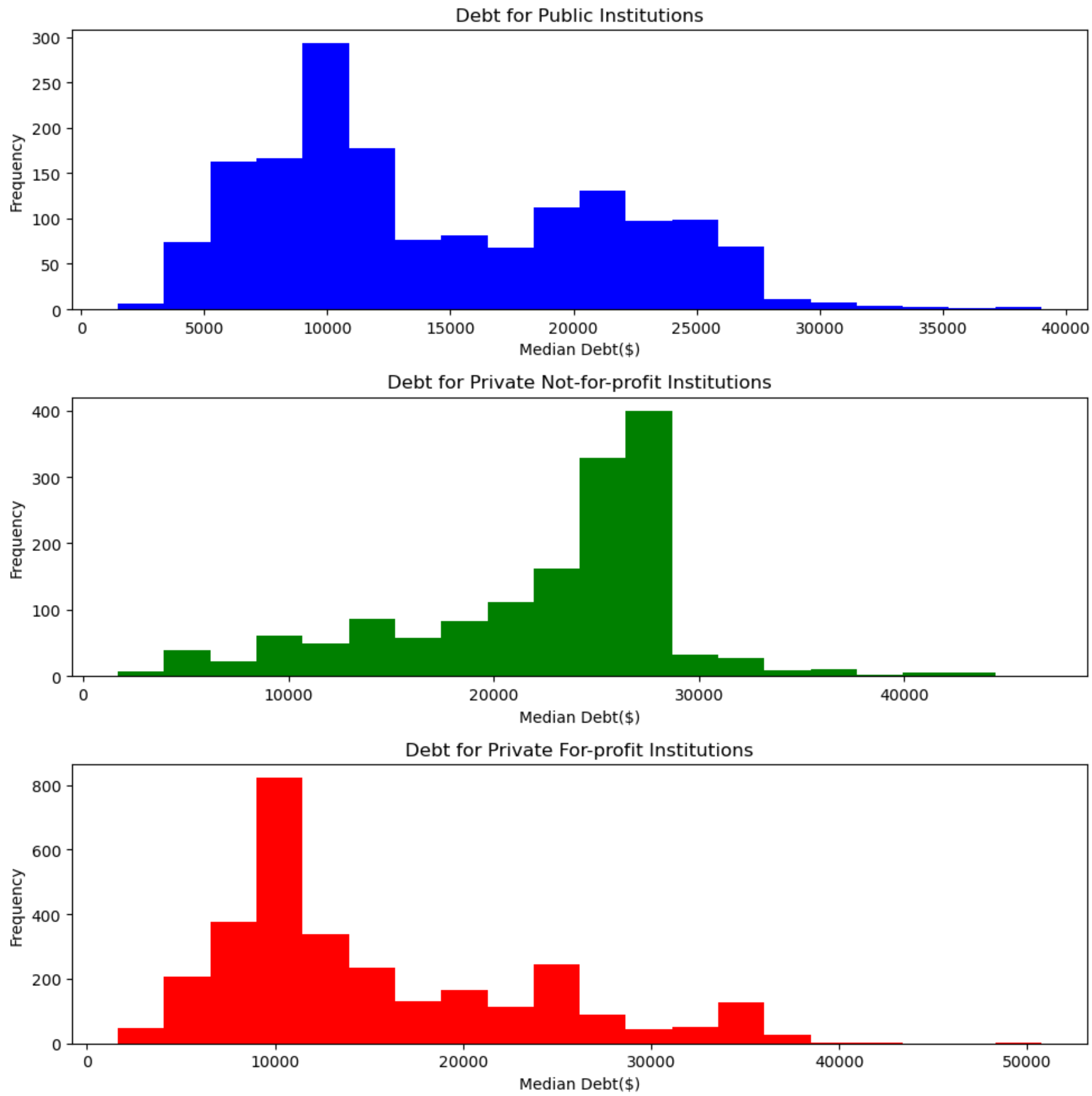
Question 3a.

Make a plot with histograms of the median debt for students who complete their degree. This plot should have three subplots, with histograms for public, private not-for-profit, and for-profit institutions.

In [8]:

```
# Filter data for different institution types
pub_inst = collegedata[(collegedata.iloc[:,16]== 1)]#public
priv_np_inst = collegedata[(collegedata.iloc[:,16]== 2)]#private nonprofit
priv_profit = collegedata[(collegedata.iloc[:,16]== 3)]#private profit institution

# Plot histograms for median debt for different institution types
plt.figure(figsize=(10, 10))

plt.subplot(311)
plt.hist(pub_inst['GRAD_DEBT_MDN'], bins=20, alpha=1, color='blue')
plt.title('Debt for Public Institutions')
plt.xlabel('Median Debt($)')
plt.ylabel('Frequency')

plt.subplot(312)
plt.hist(priv_np_inst['GRAD_DEBT_MDN'], bins=20, alpha=1, color='green')
plt.title('Debt for Private Not-for-profit Institutions')
plt.xlabel('Median Debt($)')
plt.ylabel('Frequency')
```

```
plt.subplot(313)
plt.hist(priv_profit['GRAD_DEBT_MDN'], bins=20, alpha=1, color='red')
plt.title('Debt for Private For-profit Institutions')
plt.xlabel('Median Debt($)')
plt.ylabel('Frequency')

plt.tight_layout()
plt.show()
```



Interpretation 3a.

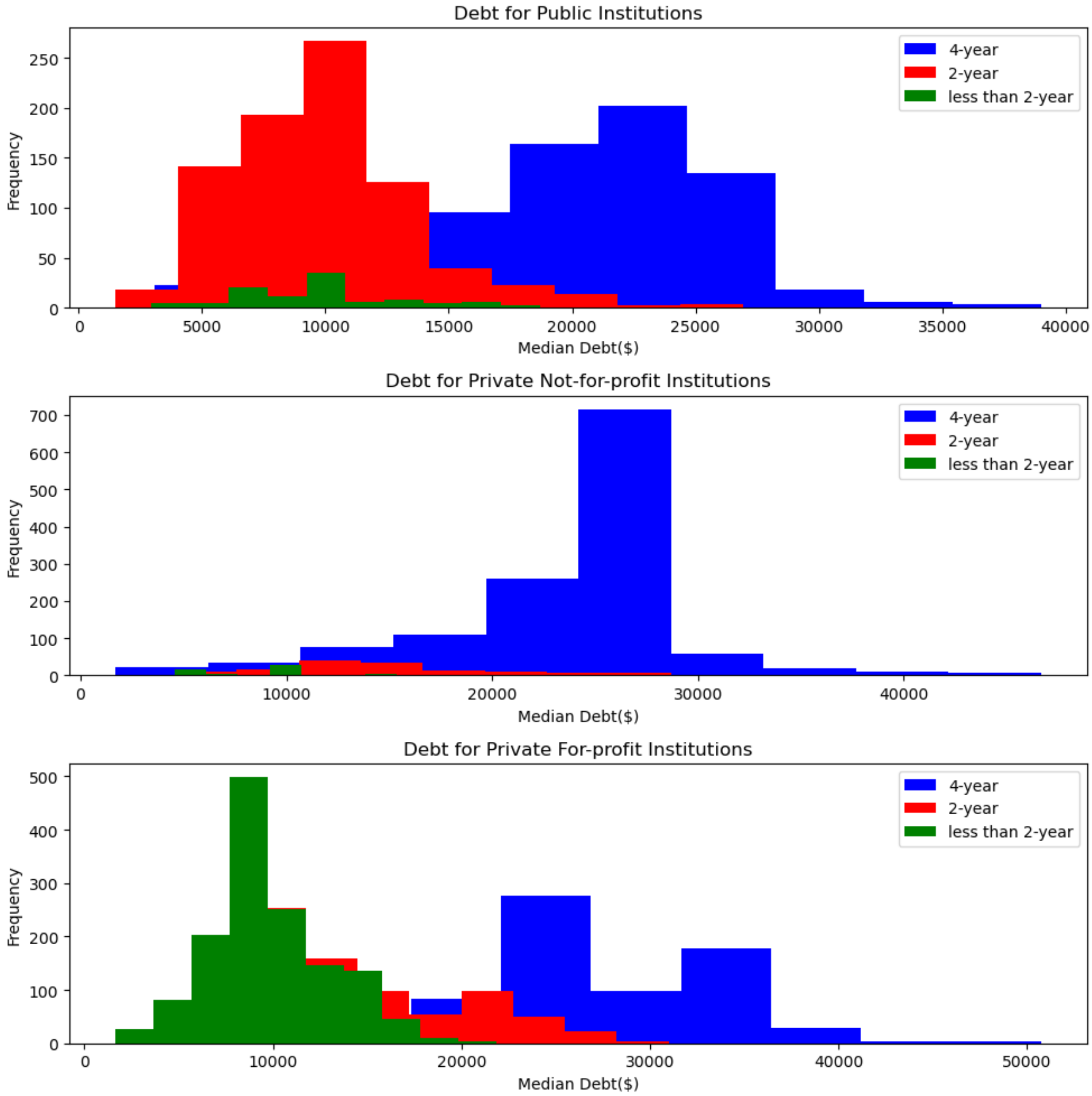
The distribution of debt for public institutions is positively skewed or skewed to the right. This means that most of these institutions have a lower median debt, while there are a few institutions with higher debt amounts. Similarly, the debt distribution for private for-profit institutions also exhibits a right-skewed pattern. In contrast, the debt distribution for private non-profit institutions is negatively skewed or skewed to the left, indicating that a larger proportion of these institutions have higher median debt amounts.

To put it simply, in the case of public and private for-profit institutions, the majority of them have lower median debt values, while a few institutions have higher debt amounts, resulting in a right-skewed distribution. On the other hand, private non-profit institutions tend to have higher median debt amounts, leading to a left-skewed distribution where a larger proportion of institutions have higher debt values.

3b

Make the same plot as part a, but within each subplot make separate colored histograms for the level of the institution (4-year, 2-year, less-than-2-year).

In [9]:
```
# Plot histograms for median debt for different institution types and their various
plt.figure(figsize=(10, 10))

plt.subplot(311)
plt.hist(pub_inst.GRAD_DEBT_MDN.loc[pub_inst.ICLEVEL==1],color="blue",label="4-year
plt.hist(pub_inst.GRAD_DEBT_MDN.loc[pub_inst.ICLEVEL==2],color="red",label="2-year
plt.hist(pub_inst.GRAD_DEBT_MDN.loc[pub_inst.ICLEVEL==3],color="green",label="less
plt.title('Debt for Public Institutions')
plt.xlabel('Median Debt($)')
plt.ylabel('Frequency')
plt.legend()

plt.subplot(312)
plt.hist(priv_np_inst.GRAD_DEBT_MDN.loc[priv_np_inst.ICLEVEL==1],color="blue",label
plt.hist(priv_np_inst.GRAD_DEBT_MDN.loc[priv_np_inst.ICLEVEL==2],color="red",label=
plt.hist(priv_np_inst.GRAD_DEBT_MDN.loc[priv_np_inst.ICLEVEL==3],color="green",lab
plt.title('Debt for Private Not-for-profit Institutions')
plt.xlabel('Median Debt($)')
plt.ylabel('Frequency')
plt.legend()


plt.subplot(313)
plt.hist(priv_profit.GRAD_DEBT_MDN.loc[priv_profit.ICLEVEL==1],color="blue",label=
plt.hist(priv_profit.GRAD_DEBT_MDN.loc[priv_profit.ICLEVEL==2],color="red",label="
plt.hist(priv_profit.GRAD_DEBT_MDN.loc[priv_profit.ICLEVEL==3],color="green",label=
plt.title('Debt for Private For-profit Institutions')
plt.xlabel('Median Debt($)')
plt.ylabel('Frequency')
plt.legend()


plt.tight_layout()
plt.show()
```



Interpretation 3b

In public institutions, the distribution of debts varies based on the duration of the program. For 4-year public institutions, the distribution is relatively symmetric, indicating a balance between lower and higher debt amounts. However, when we look at 2-year and less than 2-year public institutions, the distribution becomes skewed to the right. This suggests that a significant proportion of these institutions have lower debt amounts, while there are fewer institutions with higher debt levels.

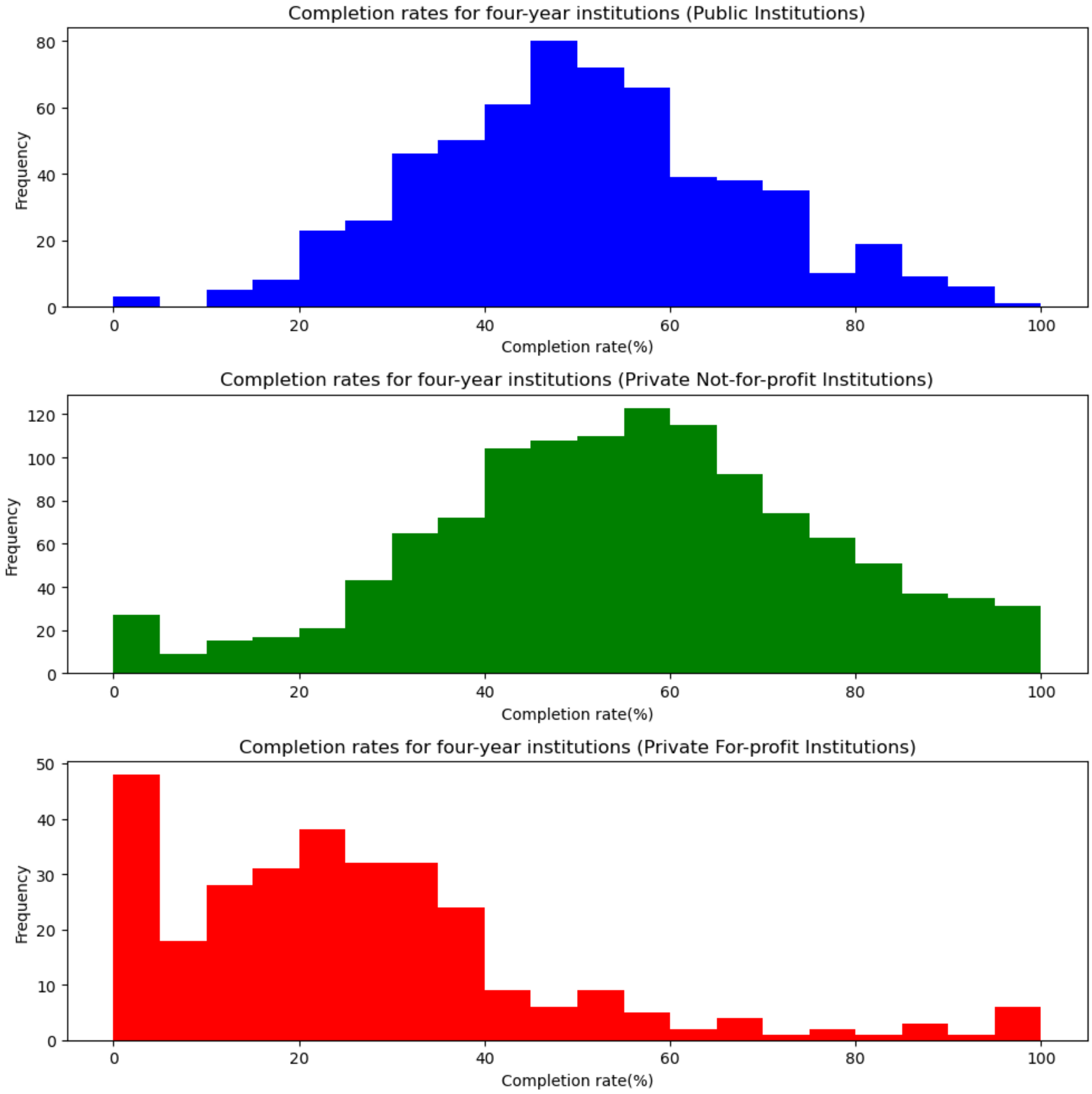
Moving on to private non-profit institutions, the distribution of debts shows a similar pattern for 4-year institutions. These institutions exhibit a relatively symmetric distribution, indicating a more balanced range of debt amounts. It is worth noting that the majority of private non-profit institutions fall under the 4-year category. However, there are also a few institutions categorized as 2-year and less than 2-year, and for these, the distribution becomes skewed to the right. This suggests that a smaller proportion of these institutions have higher debt amounts.

Lastly, in private for-profit institutions, the distribution of debts is primarily composed of institutions offering programs that are less than 2 years in duration. This category shows a right-skewed distribution, indicating that most institutions in this sector have lower debt amounts. There are also some institutions offering 2-year programs, and for these, the distribution is closer to symmetric. However, when we consider 4-year programs in private for-profit institutions, the distribution is fairly symmetric, similar to public 4-year institutions.

Question 3c

3C Make a plot showing histograms of completion rates for four-year institutions similar to part a.

In [10]:
```
#Variable C200_4 represents Completion rates for four-year institutions

plt.figure(figsize=(10, 10))

plt.subplot(311)
plt.hist(pub_inst['C200_4']*100, bins=20, alpha=1, color='blue')
plt.title('Completion rates for four-year institutions (Public Institutions)')
plt.xlabel('Completion rate(%)')
plt.ylabel('Frequency')

plt.subplot(312)
plt.hist(priv_np_inst['C200_4']*100, bins=20, alpha=1, color='green')
plt.title('Completion rates for four-year institutions (Private Not-for-profit Inst
plt.xlabel('Completion rate(%)')
plt.ylabel('Frequency')

plt.subplot(313)
plt.hist(priv_profit['C200_4']*100, bins=20, alpha=1, color='red')
plt.title('Completion rates for four-year institutions (Private For-profit Institut
plt.xlabel('Completion rate(%)')
plt.ylabel('Frequency')

plt.tight_layout()
plt.show()
```



Interpretation 3c

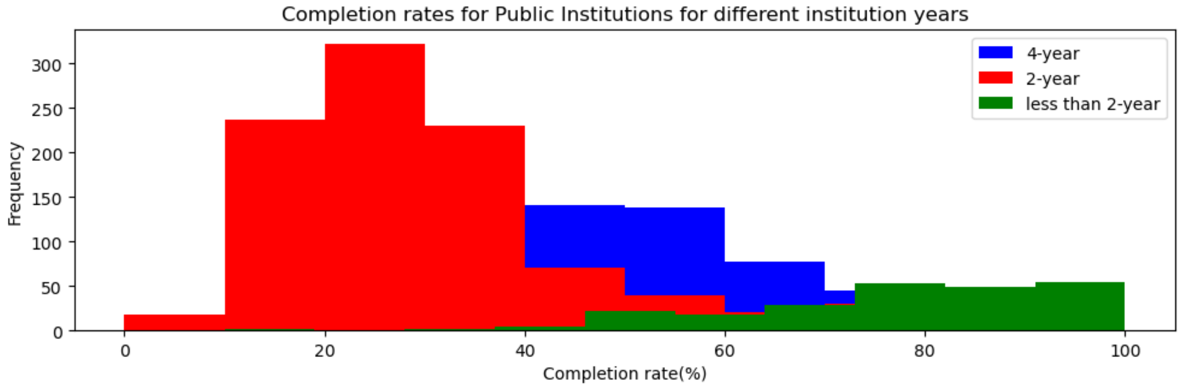
The completion rates of 4-year public and 4-year private not-for-profit institutions have a symmetrical distribution, meaning there is an equal spread of high and low completion rates. However, 4-year private for-profit institutions have a right-skewed distribution, indicating a higher concentration of lower completion rates. This suggests that completion rates vary across different types of institutions, with public and private not-for-profit institutions generally having similar completion rates while private for-profit institutions tend to have lower completion rates.

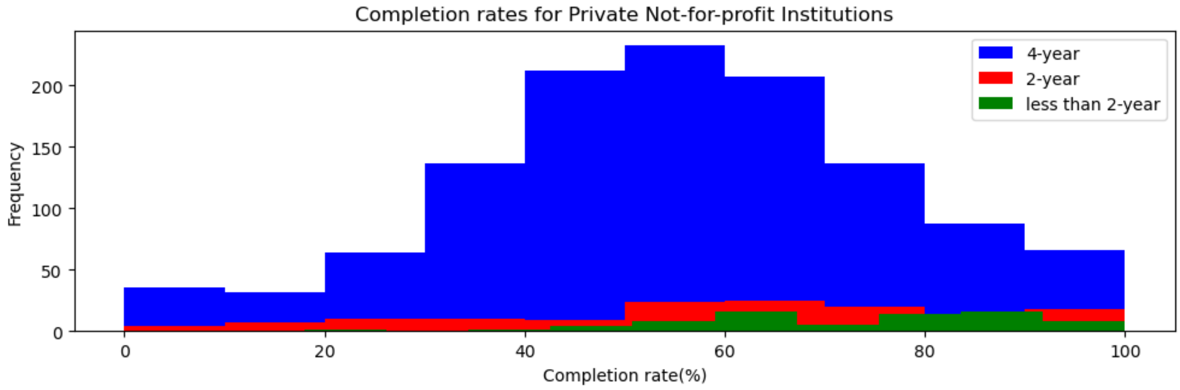
Question 3d.

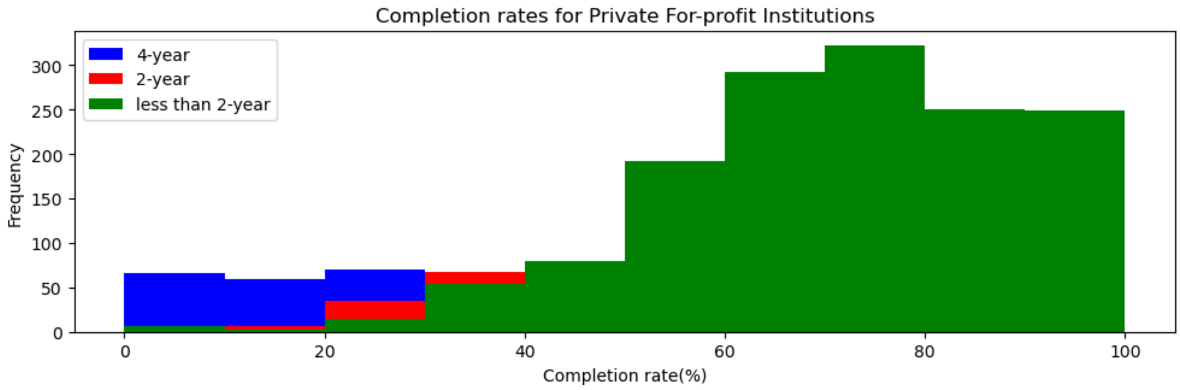
Plot histograms for completion rates for institutions for different institution types and their various ICLEVELS

In [12]:
```
##the na values were dropped here
plt.figure(figsize=(10, 10))

plt.subplot(311)
plt.hist(pub_inst.C200_4.loc[pub_inst.ICLEVEL==1].dropna().values*100 ,color="blue
plt.hist(pub_inst.C200_L4.loc[pub_inst.ICLEVEL==2].dropna().values*100 ,color="red
```

```
plt.hist(pub_inst.C200_L4.loc[pub_inst.ICLEVEL==3].dropna().values*100 ,color="gre
plt.title('Completion rates for Public Institutions for different institution year
plt.xlabel('Completion rate(%)')
plt.ylabel('Frequency')
plt.legend()

plt.subplot(312)
plt.hist(priv_np_inst.C200_4.loc[priv_np_inst.ICLEVEL==1].dropna().values*100,colo
plt.hist(priv_np_inst.C200_L4.loc[priv_np_inst.ICLEVEL==2].dropna().values*100,col
plt.hist(priv_np_inst.C200_L4.loc[priv_np_inst.ICLEVEL==3].dropna().values*100,col
plt.title('Completion rates for Private Not-for-profit Institutions')
plt.xlabel('Completion rate(%)')
plt.ylabel('Frequency')
plt.legend()


plt.subplot(313)
plt.hist(priv_profit.C200_4.loc[priv_profit.ICLEVEL==1].dropna().values*100 ,color
plt.hist(priv_profit.C200_L4.loc[priv_profit.ICLEVEL==2].dropna().values*100 ,colo
plt.hist(priv_profit.C200_L4.loc[priv_profit.ICLEVEL==3].dropna().values*100 ,colo
plt.title('Completion rates for Private For-profit Institutions')
plt.xlabel('Completion rate(%)')
plt.ylabel('Frequency')
plt.legend()


plt.tight_layout()
plt.show()
```







Question 3e

Make table with the average of median debt across institutions and the average completion rate versus both type (public, private not-for-profit, and for-profit institutions) and level (4 year, 2 year, and less than 2 year) of institution. You can calculate the values using groupby and the mean function. An example table is given below. This link shows how to make a table in Jupyter Notebook: Jupyter Notebook Users Manual.ipynb (brynmawr.edu

In [21]:
```
# Calculate average median debt and completion rate across institutions
avg_debt_completion = pd.DataFrame()

# Map ICLEVEL values to levels
level_mapping = {1: '4-yr', 2: '2-yr', 3: '<2-yr'}

#add column combine two columns
pub_inst['Completion_Rate'] = pub_inst[['C200_4', 'C200_L4']].mean(axis=1)
priv_np_inst['Completion_Rate'] = priv_np_inst[['C200_4', 'C200_L4']].mean(axis=1)
priv_profit['Completion_Rate'] = priv_profit[['C200_4', 'C200_L4']].mean(axis=1)


# Public Institutions
pub_inst['Level'] = pub_inst['ICLEVEL'].map(level_mapping)
pub_avg_debt = pub_inst.groupby(['Level']).mean()['GRAD_DEBT_MDN'].apply(lambda x:
```

```
pub_avg_completion = pub_inst.groupby(['Level']).mean()['Completion_Rate'].apply(la
pub_data = pd.DataFrame({'Type of Institution': 'Public', 'Level': pub_avg_debt.in

# Private Not-for-profit Institutions
priv_np_inst['Level'] = priv_np_inst['ICLEVEL'].map(level_mapping)
priv_np_avg_debt = priv_np_inst.groupby(['Level']).mean()['GRAD_DEBT_MDN'].apply(la
priv_np_avg_completion = priv_np_inst.groupby(['Level']).mean()['Completion_Rate']
priv_np_data = pd.DataFrame({'Type of Institution': 'Private Not-for-profit', 'Leve

# Private For-profit Institutions
priv_profit['Level'] = priv_profit['ICLEVEL'].map(level_mapping)
priv_profit_avg_debt = priv_profit.groupby(['Level']).mean()['GRAD_DEBT_MDN'].apply
priv_profit_avg_completion = priv_profit.groupby(['Level']).mean()['Completion_Rate

priv_profit_data = pd.DataFrame({'Type of Institution':'Private For-profit', 'Level

# Concatenate the dataframes
avg_debt_completion = pd.concat([pub_data, priv_np_data, priv_profit_data], ignore_


# Display the table
#print(avg_debt_completion)
```

```
C:\Users\Mary Akinde\AppData\Local\Temp\ipykernel_22236\42093827.py:15: FutureWarn
ing: The default value of numeric_only in DataFrameGroupBy.mean is deprecated. In
a future version, numeric_only will default to False. Either specify numeric_only
or select only columns which should be valid for the function.
  pub_avg_debt = pub_inst.groupby(['Level']).mean()['GRAD_DEBT_MDN'].apply(lambda
x: round(x, 0))
C:\Users\Mary Akinde\AppData\Local\Temp\ipykernel_22236\42093827.py:16: FutureWarn
ing: The default value of numeric_only in DataFrameGroupBy.mean is deprecated. In
a future version, numeric_only will default to False. Either specify numeric_only
or select only columns which should be valid for the function.
  pub_avg_completion = pub_inst.groupby(['Level']).mean()['Completion_Rate'].apply
(lambda x: round(x, 2)).replace(np.nan, 0)
C:\Users\Mary Akinde\AppData\Local\Temp\ipykernel_22236\42093827.py:21: FutureWarn
ing: The default value of numeric_only in DataFrameGroupBy.mean is deprecated. In
a future version, numeric_only will default to False. Either specify numeric_only
or select only columns which should be valid for the function.
  priv_np_avg_debt = priv_np_inst.groupby(['Level']).mean()['GRAD_DEBT_MDN'].apply
(lambda x: round(x, 0))
C:\Users\Mary Akinde\AppData\Local\Temp\ipykernel_22236\42093827.py:22: FutureWarn
ing: The default value of numeric_only in DataFrameGroupBy.mean is deprecated. In
a future version, numeric_only will default to False. Either specify numeric_only
or select only columns which should be valid for the function.
  priv_np_avg_completion = priv_np_inst.groupby(['Level']).mean()['Completion_Rat
e'].apply(lambda x: round(x, 2)).replace(np.nan, 0)
C:\Users\Mary Akinde\AppData\Local\Temp\ipykernel_22236\42093827.py:27: FutureWarn
ing: The default value of numeric_only in DataFrameGroupBy.mean is deprecated. In
a future version, numeric_only will default to False. Either specify numeric_only
or select only columns which should be valid for the function.
  priv_profit_avg_debt = priv_profit.groupby(['Level']).mean()['GRAD_DEBT_MDN'].ap
ply(lambda x: round(x, 0))
C:\Users\Mary Akinde\AppData\Local\Temp\ipykernel_22236\42093827.py:28: FutureWarn
ing: The default value of numeric_only in DataFrameGroupBy.mean is deprecated. In
a future version, numeric_only will default to False. Either specify numeric_only
or select only columns which should be valid for the function.
  priv_profit_avg_completion = priv_profit.groupby(['Level']).mean()['Completion_R
ate'].apply(lambda x: round(x, 2)).replace(np.nan, 0)
```

Type of Institution	Median Debt	Completion Rate
4-yr,Public	$20,244	0.50

Type of Institution	Median Debt	Completion Rate
4-yr,Not-for-profit	$23,755	0.55
4-yr,For-profit	$25,951	0.26
2-yr,Public	$9,752	0.32
2-yr,Not-for-profit	$15,243	0.60
2-yr,For-profit	$13,929	0.63
<2-yr,Public	$9,659	0.77
<2-yr,Not-for-profit	$8,889	0.73
<2-yr,For-profit	$9,826	0.72

Question 3f.

Discuss your findings.

Findings

The finding provides information on the median debt and completion rates for different types and levels of institutions.

For public institutions, the median debt varies across levels. For 2-year public institutions, the median debt is $9,752, while for 4-year public institutions, it is higher at \$20,245. The completion rates for these institutions also differ, with 2-year public institutions having a completion rate of 0.32, and 4-year public institutions having a higher completion rate of 0.50. Additionally, for less than 2-year public institutions, the median debt is $9,659, and the completion rate is 0.77.

Private not-for-profit institutions also exhibit variations in median debt and completion rates by level. For 2-year private not-for-profit institutions, the median debt is $15,244, and the completion rate is 0.60. In comparison, 4-year private not-for-profit institutions have a higher median debt of \$23,755 and a slightly lower completion rate of 0.55. For less than 2-year private not-for-profit institutions, the median debt is $8,889, and the completion rate is 0.73.

Private for-profit institutions also show differences in median debt and completion rates across levels. 2-year private for-profit institutions have a median debt of $13,929 and a completion rate of 0.63. 4-year private for-profit institutions have a higher median debt of \$25,951, but a lower completion rate of 0.26. Less than 2-year private for-profit institutions have a median debt of $9,826 and a completion rate of 0.72.

These findings indicate that there are variations in median debt and completion rates based on both the type (public, private not-for-profit, private for-profit) and level (2-year, 4-year, less than 2-year) of institutions. It suggests that students in different types and levels of institutions may have different financial experiences and outcomes. Students in less than 2-year institutions are more likely to finish their education process, this may be due to the shorter time involved and have the lowest debt across all institution type.

Question 3g.

Use the 2007-08 data to calculate the ratio of student loan debt to 10-year earnings. Make a table showing the mean of this ratio for each type of institution. Discuss your findings.

In [22]:
```
# Load the college data from the CSV file
collegedata0708 = pd.read_csv("C:/Users/Mary Akinde/Documents/Modern Statistics/MEI

# Define control meanings
control_meaning = {1: 'Public', 2: 'Private Nonprofit', 3: 'Private For-Profit'}

# Calculate the ratio by dividing the student loan debt (GRAD_DEBT_MDN) by the 10-y
collegedata0708['Debt_to_Earnings_Ratio'] = collegedata0708['GRAD_DEBT_MDN'] / col

# Calculate the mean ratio for each type of institution (CONTROL) using the groupby
mean_ratio_by_institution = collegedata0708.groupby(collegedata0708['CONTROL'].map

# Create a table to display the mean ratio for each type of institution
table = pd.DataFrame({'Type of Institution': mean_ratio_by_institution.index, 'Mea

#print(table)
```

```
C:\Users\Mary Akinde\AppData\Local\Temp\ipykernel_22236\568271567.py:3: DtypeWarni
ng: Columns (1) have mixed types. Specify dtype option on import or set low_memory
=False.
  collegedata0708 = pd.read_csv("C:/Users/Mary Akinde/Documents/Modern Statistics/
MERGED2007_08_PP.csv", na_values=['NULL', 'PrivacySuppressed'])
```

Type of Institition	Debt-to-Earnings Ratio
Public	0.23
Private For-Profit	0.31
Private Nonprofit	0.33

Interpretation 3G

The finding shows the mean debt-to-earnings ratio for different types of institutions.

Private for-profit institutions have the highest mean debt-to-earnings ratio, with an average value of 0.31. This indicates that, on average, the amount of student loan debt incurred by individuals attending private for-profit institutions is relatively higher compared to their earnings after 10 years. It suggests that students in these institutions may face challenges in managing their debt burden.

Private nonprofit institutions have a slightly lower mean debt-to-earnings ratio of 0.33. While still relatively high, it indicates that students in private nonprofit institutions have a slightly better balance between their debt and future earnings compared to private for-profit institutions.

Public institutions, on the other hand, have the lowest mean debt-to-earnings ratio of 0.23. This implies that students in public institutions, on average, have a more favorable debt-to-earnings ratio compared to their counterparts in private institutions. It suggests that public institutions may offer more affordable education options or have better post-graduation employment prospects, allowing students to manage their debt more effectively.

Overall, these findings highlight the importance of considering the financial implications of attending different types of institutions. It indicates that students in private for-profit and private nonprofit institutions may face higher debt burdens compared to those in public institutions.